

GenCore version 5.1.4 p5 4578  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: April 8, 2003, 10:21:07 ; Search time 333.77 Seconds  
(without alignments)  
1395.105 Million cell updates/sec

Title: US-10-020-441-3  
Perfect score: 90  
Sequence: 1 VGYGRDDNDPDRKN 16

Scoring table: BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 2054640 segs, 14551402878 residues  
Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Command line parameters:  
-MODEL=frame+ p2n.model -DEV=xlh  
-Q=/cgn2\_1/USPRO/spool/US10020441/runat\_02042003\_093012\_20970/app.query.fasta\_1.462  
-DB=genbml -QFMT=faetap -SUFFIX=rge -MINMATCH=0.1 -LOOFC=0 -LOOPEXT=0  
-UNITS=bits -START=1 -END=-1 -MATRIX=biosum62 -TRANS=human40.cdi -LIST=45  
-DOCMALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=40 -MODE=LOCAL  
-OUTMT=plc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000  
-USER=US10020441 @CGN 1 1 1758 @runat\_02042003\_093012\_20970 -NCPU=6 -ICPU=3  
-NO\_XLPRY -NO\_MMAL -LARGEQUERY -NEG\_SCORES=0 -WAIT -LONGLOG -DEV.TIMEOUT=120  
-MARK.TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7  
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Genbml:\*  
1: gb\_ba:\*  
2: gb\_hcg:\*  
3: gb\_in:\*  
4: gb\_om:\*  
5: gb\_ov:\*  
6: gb\_pat:\*  
7: gb\_ph:\*  
8: gb\_pl:\*  
9: gb\_pr:\*  
10: gb\_ro:\*  
11: gb\_sts:\*  
12: gb\_ey:\*  
13: gb\_un:\*  
14: gb\_vl:\*  
15: em\_ba:\*  
16: em\_fun:\*  
17: em\_hum:\*  
18: em\_in:\*  
19: em\_mu:\*  
20: em\_om:\*  
21: em\_or:\*  
22: em\_ov:\*  
23: em\_pat:\*  
24: em\_ph:\*  
25: em\_pl:\*  
26: em\_ro:\*  
27: em\_sts:\*  
28: em\_un:\*

29: em\_vi:\*  
30: em\_hcg\_hum:\*  
31: em\_hcg\_inv:\*  
32: em\_hcg\_other:\*  
33: em\_hcg\_mus:\*  
34: em\_hcg\_pln:\*  
35: em\_hcg\_rod:\*  
36: em\_hcg\_mam:\*  
37: em\_hcg\_vrc:\*  
38: em\_ey:\*  
39: em\_hcgo\_hum:\*  
40: em\_hcgo\_mus:\*  
41: em\_hcgo\_other:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	90	100.0	1152	3	SCMCPROT
2	90	100.0	1477	3	SMHPIELAS
3	90	100.0	3045	3	SMU31768
4	87	96.7	431	3	AF510343
5	87	96.7	943	3	AF510341
6	86	95.6	419	3	AF510342
7	81	90.0	904	3	AF510339
8	81	90.0	1163	3	SMU31769
9	54	60.0	171184	2	AC121203
10	53	58.9	164620	9	AC107218
11	52	57.8	921	9	AB083612
12	52	57.8	2697	9	AF072873
13	52	57.8	2833	6	AR129165
14	52	57.8	2880	6	AR091360
15	52	57.8	2880	6	E31206
16	52	57.8	3342	9	AB012911
17	52	57.8	31808	9	AB065702
18	52	57.8	111496	9	AC069351
19	52	57.8	160197	9	AP004221
20	52	57.8	193589	9	AC025370
21	52	57.8	347253	9	AF363578
22	51	56.7	12514	1	AB012011
23	51	56.7	20777	2	LMFLCHR34_17
24	51	56.7	110000	2	LMFLCHR32_00
25	51	56.7	110000	2	LMFLCHR34_16
26	51	56.7	199000	8	AP000396S1
27	51	56.7	201587	9	CNS01DVO
28	51	56.7	346547	1	AP003012
29	50	55.6	489	8	AF342784
30	50	55.6	146970	2	AC094913
31	50	55.6	174255	2	AC129726
32	49	54.4	91276	2	AC113361
33	49	54.4	99989	2	AC096261
34	49	54.4	110000	2	AC096785_0
35	49	54.4	161188	2	AC095827
36	49	54.4	195653	2	AC108127
37	49	54.4	201070	2	AC026291
38	49	54.4	214792	2	AC102091
39	48.5	53.9	35811	8	SC9718
40	48	53.3	886	3	AF510340
41	48	53.3	7906	6	AX277935
42	48	53.3	7906	6	AX323622
43	48	53.3	7906	6	AX344771
44	48	53.3	8457	1	MSP011317
45	48	53.3	87487	8	AB015474

RESULT 1

ALIGNMENTS

SCMCPROT 1152 bp mRNA linear INV 26-APR-1993  
 LOCUS S. mansoni cercarial protease mRNA, complete cds.  
 DEFINITION J03946  
 J03946.1 GI:160947  
 VERSION cercarial protease; serine protease.  
 KEYWORDS S. mansoni (strain Puerto Rican) cercaria acetabular gland, cDNA to  
 SOURCE mRNA, clone B.  
 ORGANISM Schistosoma mansoni  
 Eukaryota; Metazoa; Platyhelminthes; Trematoda; Digenea;  
 Strigoidida; Schistosomatidae; Schistosomatidae; Schistosoma.  
 REFERENCE 1 (bases 1 to 1152)  
 NEWPORT, G.R., MCKERROW, J.H., HEDSTROM, R., PETILT, M., MCGARRIGLE, L.,  
 BARR, P.J. and AGABIAN, N.  
 Cloning of the proteinase that facilitates infection by schistosome  
 parasites  
 J. Biol. Chem. 263 (26), 13179-13184 (1988)  
 MEDLINE 88330818  
 PUBMED 3166457  
 COMMENT Draft entry and clean copy of sequence [1] kindly submitted by  
 J.H.McKerrow, 13-JUN-1998  
 FEATURES  
 source location/Qualifiers  
 1..1152  
 /organism="Schistosoma mansoni"  
 /db\_xref="taxon:6183"  
 255..1049  
 /note="cercarial protease precursor"  
 /codon\_start=1  
 /protein\_id="AA029864.1"  
 /db\_xref="GI:160948"  
 /translation="MSNMRFPVVVLTFTYCLTFERVSTWLRISGEPIVQHPAEPFPIA  
 FLTTERTCSTSLVSTRVATAGCVCSPFLVIRVSPFLNNGOQGIHOPSGVKVA  
 PGVWSCSARORPIAQTLSGDIATVMAQVNLQSGIRVISPQSDIPLPPTGV  
 FIVGVRDNDRPSRKNQGIKRGATIMCRHATNENPICVARAGNFGGLPARGDS  
 GGPLPSIQGVLVSHGVTLPNLPDIIVEYASVARMLEDFRSNI"  
 sig\_peptide 255..335  
 /note="cercarial protease signal peptide"  
 mat\_peptide 336..1046  
 /product="cercarial protease"  
 BASE COUNT 345 a 273 c 230 g 304 t  
 ORIGIN 369 bp upstream of EcoRI site.  
 Alignment Scores:  
 Pred. No.: 7.02e-05 Length: 1152  
 Score: 90.00 Matches: 16  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 100.00% Indels: 0  
 DB: 3 Gaps: 0  
 US-10-020-441-3 (1-16) x SCMCPROT (1-1152)  
 QY 1 VALGIYTYRGLYARGASPAPASPAATGAPPROSErArgLYsAsn 16  
 |||||  
 Db 741 GTTGGTTATGGAAGGATGATACGACCGTGATCCGTACGTAAGAAAT 788  
 RESULT 2  
 SMPIELIAS  
 LOCUS SMPIELIAS 1477 bp DNA linear INV 29-MAR-1996  
 DEFINITION S. mansoni elastase HPI gene.  
 ACCESSION Z70296  
 VERSION Z70296.1 GI:1240057  
 KEYWORDS elastase.  
 SOURCE Schistosoma mansoni.  
 ORGANISM Schistosoma mansoni  
 Eukaryota; Metazoa; Platyhelminthes; Trematoda; Digenea;  
 Strigoidida; Schistosomatidae; Schistosomatidae; Schistosoma.  
 REFERENCE 1 (bases 1 to 1477)  
 Sayers, J.R., Price, H.P. and Doenhoff, M.J.  
 JOURNAL Unpublished  
 REFERENCE 2 (bases 1 to 1477)  
 Sayers, J.R.  
 AUTHORS Direct Submission  
 TITLE

JOURNAL Submitted (21-MAR-1996) Jon R Sayers, Medicine and Pharmacology,  
 University of Sheffield, Royal Hallamshire Hospital, Sheffield,  
 South Yorkshire, S10 2JF, UK  
 FEATURES  
 source location/Qualifiers  
 1..1477  
 /organism="Schistosoma mansoni"  
 /db\_xref="taxon:6183"  
 16..168  
 /number=1  
 join(88..168,310..627,1235..1477)  
 /note="consensus sequence of 3 PCR clones from genomic  
 DNA"  
 /codon\_start=1  
 /product="cercarial elastase HPI"  
 /protein\_id="CAA94312.1"  
 /db\_xref="GI:1240058"  
 /translation="MCCTGSLVSTRVATAGCVCSPFLVIRVSPFLNNGOQGIH  
 PSGVKVAPGTWPSCSARORPIAQTLSGDIATVMAQVNLQSGIRVISPQSDI  
 PLPPTGVFIVGVRDNDRPSRKNQGIKRGATIMCRHATNENPICVARAGNFG  
 GLPARGDSGPLPSIQGVLVSHGVTLPNLPDIIVEYASVARMLEDFRSNI"  
 169..309  
 /number=1  
 310..627  
 /number=2  
 628..1234  
 /number=2  
 1235..1477  
 /number=3  
 BASE COUNT 347 a 261 c 404 g 465 t  
 ORIGIN  
 Alignment Scores:  
 Pred. No.: 8.76e-05 Length: 1477  
 Score: 90.00 Matches: 16  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 100.00% Indels: 0  
 DB: 3 Gaps: 0  
 US-10-020-441-3 (1-16) x SMPIELIAS (1-1477)  
 QY 1 VALGIYTYRGLYARGASPAPASPAATGAPPROSErArgLYsAsn 16  
 |||||  
 Db 562 GTTGGTTATGGAAGGATGATACGACCGTGATCCGTACGTAAGAAAT 609  
 RESULT 3  
 SMUJ1768  
 LOCUS SMUJ1768 3045 bp DNA linear INV 28-JAN-1996  
 DEFINITION Schistosoma mansoni elastase gene, 3045 bp clone, complete cds.  
 ACCESSION U31768  
 VERSION U31768  
 KEYWORDS U31768.1 GI:1103828  
 SOURCE blood fluke.  
 ORGANISM Schistosoma mansoni  
 Eukaryota; Metazoa; Platyhelminthes; Trematoda; Digenea;  
 Strigoidida; Schistosomatidae; Schistosomatidae; Schistosoma.  
 REFERENCE 1 (bases 1 to 3045)  
 Pierrot, C., Capron, A. and Khalife, J.  
 Cloning and characterization of two genes encoding Schistosoma  
 mansoni elastase  
 Mol. Biochem. Parasitol. 75 (1), 113-117 (1995)  
 JOURNAL 96362066  
 MEDLINE 8720180  
 PUBMED 8720180  
 REFERENCE 2 (bases 1 to 3045)  
 Pierrot, C.  
 AUTHORS Direct Submission  
 TITLE Submitted (18-JUL-1995) Christine Pierrot, CIBP, INSERM U 167 -  
 Institut Pasteur de Lille, 1, rue du Professeur Calmette, Lille  
 59019, France  
 FEATURES  
 source location/Qualifiers  
 1..3045  
 /organism="Schistosoma mansoni"  
 /strain="Puerto rico"



```
mRNA
/d,db_xref="taxon:6183"
join(1..1465,1588..1935,2525..2910)
/product="elastase"
5'UTR
1..1231
1153..1160
1162..1167
join(1232..1465,1588..1935,2525..2767)
/codon_start=1
/codon_end=1
/product="elastase"
/db_xref="GI:1103829"
/translation="MSNRRLVLTFTYCLTFRVSTWLRSGEPVQRTPEPIAF
LTERMTCTSLVSTRVLTAGHCVCSPLPVIRVLCFQVSPFLTRNGDQGIHQPS
GVKVAFGVMPSCMSARGRPIAQLTSLAIVLQVNIQSGITVSLPQASDIPF
PGTGVFIYVGRDNDPDRNGSLIKKELVGRATIMECRHATNGNPICVKAGCN
FGOLPAPDSSGPIPLSPQCPVLCVSHGVTLPNLPDIIIVEYASVARNLDFVRNI"
3'UTR
2768..2910
polya_signal
2886..2891
polya_signal
2906..2910
BASE COUNT 851 a 579 c 691 g 924 t
ORIGIN

Alignment Scores:
Pred. No.: 0.000166 Length: 3045
Score: 90.00 Matches: 16
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 3 Gaps: 0

US-10-020-441-3 (1-16) x SMUJ1768 (1-3045)
Cy 1 ValGlyTYrGlyArGAspAspAspArGAspProSerArGlyAsn 16
Db 1855 GTTGGTATTGAGAGGATGATACGACCGTGATCCGTCGTGAATAAT 1902

RESULT 4
AF510343 431 bp mRNA linear INV 01-JUL-2002
LOCUS Schistosomium douthitti elastase 1a (CE-1a) mRNA, partial cds.
DEFINITION AF510343
ACCESSION AF510343.1 GI:21217538
VERSION AF510343.1
KEYWORDS Schistosomium douthitti.
SOURCE Schistosomium douthitti.
ORGANISM Eukaryota; Metazoa; Platyhelminthes; Trematoda; Digenea;
Strigoidida; Schistosomatidae; Schistosomatidae; Schistosomium.
REFERENCE 1 (bases 1 to 431)
AUTHORS Salter,J.P., Choe,Y., Albrecht,H., Franklin,C., Lim,K.C.,
Craig,K.S. and McKerrrow,J.H.
Cercarial elastase is encoded by a functionally conserved gene
family across multiple species of schistosomes
J. Biol. Chem. 277 (27), 24618-24624 (2002)
TITLE Cercarial elastase is encoded by a functionally conserved gene
family across multiple species of schistosomes
JOURNAL J. Biol. Chem. 277 (27), 24618-24624 (2002)
MEDLINE 22086216
PUBMED 11986325
REFERENCE 2 (bases 1 to 431)
AUTHORS Salter,J.P., Choe,Y., Albrecht,H., Franklin,C., Lim,K.-C., Craig,C.
and McKerrrow,J.H.
Direct Submission
Submitted (08-MAY-2002) Pathology, University of California - San
Francisco, 513 Parnassus, Box 0511, San Francisco, CA 94143, USA
FEATURES
source
1..431
/organism="Schistosomium douthitti"
/db_xref="taxon:38741"
<1..>431
/gene="CE-1a"
/note="SDCE-1a"
<1..>431
/gene="CE-1a"
/feature="SDCE-1a"
<1..>431
/gene="CE-1a"
/feature="protease"
/codon_start=1
/product="elastase 1a"
BASE COUNT 263 a 220 c 210 g 250 t
```

```
/protein_id="AA043945.1"
/db_xref="GI:21217539"
/translation="DOGIHQPSGVVAVPEYMPSCASRRRRVKTLSGFDIAIVL
LAEMVNIQTGIXVLSLPQDPIPARGPVPIVYGRDNDPDRNRNGILKKRATV
MESKSTTGNDICVGRVIFRQDIPAPDSSGPIPLSPQNSIS"
BASE COUNT 117 a 108 c 109 g 97 t
ORIGIN

Alignment Scores:
Pred. No.: 8.52e-05 Length: 431
Score: 87.00 Matches: 15
Percent Similarity: 100.00% Conservative: 1
Best Local Similarity: 93.75% Mismatches: 0
Query Match: 96.67% Indels: 0
DB: 3 Gaps: 0

US-10-020-441-3 (1-16) x AF510343 (1-431)
Cy 1 ValGlyTYrGlyArGAspAspAspArGAspProSerArGlyAsn 16
Db 226 GTTGGTATTGAGAGGATGATACGACCGTGATCCGTCACGAGAAAT 273

RESULT 5
AF510341 943 bp mRNA linear INV 01-JUL-2002
LOCUS Schistosoma haematobium elastase 1a (CE-1a) mRNA, complete cds.
DEFINITION AF510341
ACCESSION AF510341
VERSION AF510341.1 GI:21217534
KEYWORDS Schistosoma haematobium.
SOURCE Schistosoma haematobium.
ORGANISM Eukaryota; Metazoa; Platyhelminthes; Trematoda; Digenea;
Strigoidida; Schistosomatidae; Schistosomatidae; Schistosoma.
REFERENCE 1 (bases 1 to 943)
AUTHORS Salter,J.P., Choe,Y., Albrecht,H., Franklin,C., Lim,K.C.,
Craig,K.S. and McKerrrow,J.H.
Cercarial elastase is encoded by a functionally conserved gene
family across multiple species of schistosomes
J. Biol. Chem. 277 (27), 24618-24624 (2002)
TITLE Cercarial elastase is encoded by a functionally conserved gene
family across multiple species of schistosomes
JOURNAL J. Biol. Chem. 277 (27), 24618-24624 (2002)
MEDLINE 22086216
PUBMED 11986325
REFERENCE 2 (bases 1 to 943)
AUTHORS Salter,J.P., Choe,Y., Albrecht,H., Franklin,C., Lim,K.-C., Craig,C.
and McKerrrow,J.H.
Direct Submission
Submitted (08-MAY-2002) Pathology, University of California - San
Francisco, 513 Parnassus, Box 0511, San Francisco, CA 94143, USA
FEATURES
source
1..943
/organism="Schistosoma haematobium"
/db_xref="taxon:6185"
1..943
/gene="CE-1a"
/feature="SDCE-1a"
1..117
/feature="SDCE-1a"
18..809
/gene="CE-1a"
/feature="protease"
/codon_start=1
/product="elastase 1a"
/protein_id="AA043943.1"
/db_xref="GI:21217535"
/translation="MLNRRLVLTFTYCLTFRVSTWLRSGEPVQRTPEPIAF
LTERMTCTSLVSTRVLTAGHCVCSPLPVIRVLCFQVSPFLTRNGDQGIHQPSGVVAVP
GVKVAFGVMPSCMSARGRPIAQLTSLAIVLQVNIQSGITVSLPQDPIPARGPVPIV
YGRDNDPDRNRNGSLIKKELVGRATIMECRHATNGNPICVKAGCNFGOLPAPDSSG
PIPLSPQCPVLCVSHGVTLPNLPDIIIVEYASVARNLDFVRNI"
810..943
/gene="CE-1a"
/feature="protease"
/codon_start=1
/product="elastase 1a"
BASE COUNT 263 a 220 c 210 g 250 t
```

## ORIGIN

## Alignment Scores:

Pred. No.:	0.000171	Length:	943
Score:	87.00	Matches:	15
Percent Similarity:	100.00%	Conservative:	1
Best Local Similarity:	93.75%	Mismatches:	0
Query Match:	96.67%	Indels:	0
DB:	3	Gaps:	0

US-10-020-441-3 (1-16) x AF510341 (1-943)

Qy 1 ValGlyTYrGlyArgAspAspAspAspArgAspProSerArgLyAsn 16

Db 501 GTTGTTATGGAAGGATGATACGACCGTATCCGTACCGAGAAT 548

RESULT 6  
AF510342 419 bp mRNA linear INV 01-JUL-2002LOCUS Schistosoma haematobium elastase 1b (CE-1b) mRNA, partial cds.  
DEFINITION AF510342  
ACCESSION AF510342.1 GI:21217536

KEYWORDS Schistosoma haematobium.

SOURCE Schistosoma haematobium.

REFERENCE Salter,J.P., Choe,Y., Albrecht,H., Franklin,C., Lim,K.C.,  
AUTHORS Craik,C.S. and McKerrow,J.H.  
TITLE Cercarial elastase is encoded by a functionally conserved gene  
JOURNAL family across multiple species of schistosomes  
MEDLINE J. Biol. Chem. 277 (27), 24618-24624 (2002)  
PUBMED 11986325REFERENCE 2 (bases 1 to 419)  
AUTHORS Salter,J.P., Choe,Y., Albrecht,H., Franklin,C., Lim,K.-C., Craik,C.  
TITLE Direct Submission  
JOURNAL Submitted (08-MAY-2002) Pathology, University of California - San  
JOURNML Francisco, Box 0511, San Francisco, CA 94143, USAFEATURES  
source Location/Qualifiers  
1..419/organism="Schistosoma haematobium"  
/db\_xref="taxon:6185"<1..>419  
/gene="CE-1b"<1..>419  
/note="ShCE-1b"

/gene="CE-1b"

/note="protease"

/codon\_start=1  
/product="elastase 1b"

/protein\_id="AA043944.1"

/db\_xref="GI:21217537"

/translation="DOGGIHQPSGVVAPGPMSCMARQRPRIKOTLSGFDAIVS  
LAQVNLQGTGIRVLSLPQPTDIPRGTVGIVGARDNDNDPQKNGSILKKGRATI  
MECHATNGDPICTVSGNFGCLPARGSGEPLPSP"

BASE COUNT 118 a 104 c 102 g 95 t

ORIGIN

Alignment Scores:

Pred. No.:	0.000118	Length:	419
Score:	86.00	Matches:	15
Percent Similarity:	93.75%	Conservative:	0
Best Local Similarity:	93.75%	Mismatches:	1
Query Match:	95.56%	Indels:	0
DB:	3	Gaps:	0

US-10-020-441-3 (1-16) x AF510342 (1-419)

Qy 1 ValGlyTYrGlyArgAspAspAspAspArgAspProSerArgLyAsn 16

Db 501 GTTGTTATGGAAGGATGATACGACCGTATCCGTACCGAGAAT 548

RESULT 6  
AF510342 419 bp mRNA linear INV 01-JUL-2002LOCUS Schistosoma haematobium elastase 1b (CE-1b) mRNA, partial cds.  
DEFINITION AF510342  
ACCESSION AF510342.1 GI:21217536

KEYWORDS Schistosoma haematobium.

SOURCE Schistosoma haematobium.

Db 226 GTTGTTATGGAAGGATGATACGACCGTATCCACAGCTAAAT 273

RESULT 7  
AF510339 904 bp mRNA linear INV 01-JUL-2002LOCUS Schistosoma mansoni elastase 2a (CE-2a) mRNA, complete cds.  
DEFINITION AF510339  
ACCESSION AF510339  
VERSION AF510339.1 GI:21217530

KEYWORDS Schistosoma mansoni.

SOURCE Schistosoma mansoni.

REFERENCE Salter,J.P., Choe,Y., Albrecht,H., Franklin,C., Lim,K.C.,  
AUTHORS Craik,C.S. and McKerrow,J.H.  
TITLE Cercarial elastase is encoded by a functionally conserved gene  
JOURNAL family across multiple species of schistosomes  
MEDLINE J. Biol. Chem. 277 (27), 24618-24624 (2002)  
PUBMED 11986325REFERENCE 2 (bases 1 to 904)  
AUTHORS Salter,J.P., Choe,Y., Albrecht,H., Franklin,C., Lim,K.-C., Craik,C.  
TITLE Direct Submission  
JOURNAL Submitted (08-MAY-2002) Pathology, University of California - San  
JOURNML Francisco, Box 0511, San Francisco, CA 94143, USAFEATURES  
source Location/Qualifiers  
1..904/organism="Schistosoma mansoni"  
/db\_xref="taxon:6183"1..904  
/gene="CE-2a"1..792  
/note="SMCE-2a"

/gene="CE-2a"

/note="protease"

/codon\_start=1  
/product="elastase 2a"

/protein\_id="AA043941.1"

/db\_xref="GI:21217531"

/translation="WLNKRTFLMYTLFTTCITPRVSTWLVRKPEVQDRTEPYIAF  
VTRRTMCTGSLVSTRAVLTAAGHCVCSPMPVQVSLTLNNGDQGIHQPSGVVAP  
EYMPSCYASBQRRIKOTLSGFDAIVSIPQASDIPPTGVDF  
IVGGRDNDNDPESRAGILKKGRATVMECKSTNPICVQAAVFGQITAPDSDG  
GPILRSPQGPVLGVSHVLTLSNRLDVLEVASVARNLGFVSSNI"

BASE COUNT 239 a 200 c 224 g 241 t

ORIGIN

polyA\_signal  
/gene="CE-2a"

Alignment Scores:

Pred. No.:	0.00138	Length:	904
Score:	81.00	Matches:	14
Percent Similarity:	100.00%	Conservative:	1
Best Local Similarity:	93.33%	Mismatches:	0
Query Match:	90.00%	Indels:	0
DB:	3	Gaps:	0

US-10-020-441-3 (1-16) x AF510339 (1-904)

Qy 1 ValGlyTYrGlyArgAspAspAspAspArgAspProSerArgLyAsn 15

Db 484 GTTGTTATGGAAGGATGATACGACCGTATCCGTACCGAGA 528

RESULT 8  
SMU31769 1163 bp DNA linear INV 28-JAN-1996LOCUS Schistosoma mansoni elastase gene, 1163 bp clone, complete cds.  
DEFINITION U31769  
ACCESSION U31769.1 GI:1103830

KEYWORDS blood fluke.

SOURCE

ORGANISM  
Eukaryota: Metazoa: Platyhelminthes: Trematoda: Digenea: Strigeidae, Schistosomatoidea: Schistosomatidae, Schistosoma.

REFERENCE  
1 (bases 1 to 1163)  
Pierrot, C., Capron, A. and Khalife, J.  
Cloning and characterization of two genes encoding Schistosoma mansoni elastase

JOURNAL  
Mol. Biochem. Parasitol. 75 (1), 113-117 (1995)

MEDLINE  
96362066

PUBMED  
8720180

REFERENCE  
2 (bases 1 to 1163)  
Pierrot, C.  
Direct Submission  
Submitted (18-JUL-1995) Christine Pierrot, CIBP, INSERM U167 - Institut Pasteur de Lille, 1, rue du Professeur Calmette, Lille, 59019 France

FEATURES  
source  
1..1163  
/organism="Schistosoma mansoni"  
/strain="Puerto rico"  
/db\_xref="taxon:6183"  
join<1..234,360..668,933..21163)  
join<1..234,360..668,933..1163)  
/codon\_start=1  
/product="elastase"  
/protein\_id="AAC46968.1"  
/db\_xref="GI:1103831"  
/translation="MSNRMFLVYVTLFYICLTFKRVSTWLRSDPVQKHTPEFPIAY  
IASKSGTGLSVTRVYTAGHCVCPEPVYIKVYVTELTNRNGQSGIHQSGVKV  
APEVNPSTASRORRIRIQTLSGFAIAAMCAWVNLQSGIRVIGLPAOSDIPPGTGT  
VFVIGRGNDDRDPSRPRAGGRATVETACREKHTVNPICVYAGAPRNSGDLDPGSGSG  
LIPSGPVLGVSHVGHVTLSHLPERVSVARMLNFRSNI"

BASE COUNT  
311 a 226 c 275 g 351 t

ORIGIN

Alignment Scores:  
Pred. No.:  
Score: 0.00173 Length: 1163  
Percent Similarity: 100.00% Matches: 14  
Best Local Similarity: 93.33% Mismatches: 1  
Query Match: 90.00% Indels: 0  
DB: 3 Gaps: 0

US-10-020-441-3 (1-16) x SMUJ1769 (1-1163)

QY 1 ValGIYTYRGIVARGAASPAPsapaSPARGAAPProSeArzLyLys 15  
GGTGGTTATGGAAGGATGATACGACCGTGATCCGTCCGTAAGA 659

Db 615

RESULT 9  
AC121203/c

LOCUS  
AC121203/c

DEFINITION  
Rattus norvegicus clone CH230-120B22, \*\*\* SEQUENCING IN PROGRESS

VERSION  
AC121203

KEYWORDS  
HTG; HTGS; PHASE1.

SOURCE  
Rattus norvegicus

ORGANISM  
Eukaryota: Metazoa: Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.

REFERENCE  
1 (bases 1 to 171184)  
Munry, D.M., Adams, C., Adio-Oduola, B., Ali-Osman, F.R., Allen, C., Albrooks, S.L., Amaralunga, H.C., Are, J.R., Ayale, M., Banks, T., Barbarta, J., Benton, J., Blinze, K., Blankenburg, K., Bonnin, D., Bouck, J., Bowde, S., Brieva, M., Brown, E., Brown, M., Bryant, N.P., Bunay, C., Burke, P., Burkett, C., Buttrill, K.L., Byrd, N.C., Caron, T.F., Carter, M., Cavazos, S.R., Chacko, J., Chavez, D., Chen, G., Chen, R., Chen, Z., Chowdhury, I., Christopoulos, C., Cleveland, C.D., Cox, C., Coyle, M.D., Dathorne, S.R., David, R., Delaney, M.L., Davis, C., Davy-Carroll, L., Decker, D.A., Delaney, K.R., Delgado, O., Dem, A.L., Ding, Y., Dinn, H.H.,

Douthett, K.J., Draper, H., Dugan-Rocha, S., Durbin, K.J.,  
 Barnhart, C., Edgar, D., Edwards, C.C., Elhaj, C., Escotto, M.,  
 Falls, T., Ferraguto, D., Flagg, N., Ford, J., Foster, P., Frantz, P.,  
 Gabis, A., Gao, J., Garcia, A., Garner, T., Garza, N., Gill, R.,  
 Gorrell, J.H., Guevara, W., Gunaratne, P., Hale, S., Hamilton, K.,  
 Harris, C., Harris, K., Hart, M., Havlak, P., Hawes, A., Hernandez, J.,  
 Hernandez, O., Hodgson, A., Hogues, M., Holloway, C., Hollins, B.,  
 Homs, F., Howard, S., Huber, J., Hulik, S., Hume, J., Jackson, L.E.,  
 Jacobson, B., Jia, Y., Johnson, R., Jolivet, S., Joudah, S.,  
 Karlsson, E., Kelly, S., Khan, U., King, L., Korvah, J., Kovar, C.,  
 Kratovic, J., Kuzschi, A., Landry, N., Leal, B., Lewis, L.C., Lewis, L.,  
 Li, Z., Lichtarge, O., Lieu, C., Liu, J., Liu, J., Liu, W., Louie, H.,  
 Lozada, R.J., Lu, X., Lucier, A., Lucier, R., Luna, R., Ma, J.,  
 Maheshwari, M., Mapue, P., Martin, R., Martindale, A., Martinez, E.,  
 Massey, E., Maunthly, E., McLeod, M.P., Meador, M., Mei, G., Metzger, M.,  
 Miner, G., Miner, Z., Mitchell, T., Mohabbat, K., Morgan, M., Morris, S.,  
 Moser, M., Neal, D., Newton, J., Newton, N., Nguyen, A., Nguyen, N.,  
 Nguyen, N., Nickerson, E., Nwokoko, S., Ogun, M., Okunolu, G.,  
 Oragunye, N., Oviedo, R., Pace, A., Payton, B., Peery, J., Perez, L.,  
 Peters, L., Pichens, R., Primus, E., Pul, L., Quiles, M., Ren, Y.,  
 Rivers, M., Rojas, A., Rojibokan, I., Rolfe, M., Ruiz, S., Savary, G.,  
 Scherer, S., Scott, G., Shen, H., Shochet, N., Slason, I.,  
 Segregan, E., Senaite, T., Sparks, A., Stanley, H., Stone, H.,  
 Sutton, A., Svatek, A., Taber, P., Tameris, A., Tameris, K., Tang, H.,  
 Tansey, J., Taylor, C., Taylor, T., Telford, B., Thomas, N., Thomas, S.,  
 Usmani, K., Vazquez, L., Vera, V., Villalon, D., Vinson, R., Wang, Q.,  
 Wang, S., Ward-Moore, S., Warren, R., Washington, C., Watlington, S.,  
 Williams, G., Williamson, A., Wiczek, R., Wooten, S., Worley, K.,  
 Wu, C., Wu, Y., Zhou, J., Zorrilla, S., Nelson, D.,  
 Weinstein, G. and Gibbs, R.  
 Direct Submission  
 Unpublished  
 2 (bases 1 to 171184)  
 Worley, K.C.  
 Direct Submission  
 Submitted (16-MAY-2002) Human Genome Sequencing Center, Department  
 of Molecular and Human Genetics, Baylor College of Medicine, One  
 Baylor Plaza, Houston, TX 77030, USA  
 3 (bases 1 to 171184)  
 Worley, K.C.  
 Direct Submission  
 Submitted (23-JUL-2002) Human Genome Sequencing Center, Department  
 of Molecular and Human Genetics, Baylor College of Medicine, One  
 Baylor Plaza, Houston, TX 77030, USA  
 On Jul 19, 2002 this sequence version replaced gi:20806233.  
 ----- Genome Center  
 Center: Baylor College of Medicine  
 Center code: BCM  
 Web site: <http://www.hgsc.bcm.tmc.edu/>  
 Contact: [hgsc-help@bcm.tmc.edu](mailto:hgsc-help@bcm.tmc.edu)  
 ----- Project Information  
 Center project name: GKUL  
 Center clone name: CH230-120R22  
 ----- Summary Statistics  
 Sequencing vector: Plasmid;  
 Chemistry: Dye-terminator Big Dye; 100% of reads  
 Assembly program: Phrap; version 0.990329  
 Consensus quality: 82667 bases at least Q40  
 Consensus quality: 89424 bases at least Q30  
 Consensus quality: 95128 bases at least Q20  
 -----  
 \* NOTE: Estimated insert size may differ from sequence length  
 \* (see [http://www.hgsc.bcm.tmc.edu/docs/genbankdraft\\_data.html](http://www.hgsc.bcm.tmc.edu/docs/genbankdraft_data.html))  
 \* NOTE: This is a 'working draft' sequence. It currently  
 \* consists of 83 contigs. The true order of the pieces  
 \* is not known and their order in this sequence record is  
 \* arbitrary. Gaps between the contigs are represented as  
 \* runs of N, but the exact sizes of the gaps are unknown.  
 \* This record will be updated with the finished sequence  
 \* as soon as it is available and the accession number will  
 \* be preserved.  
 \*  
 \* 1 1176: contig of 1176 bp in length  
 \* 1177 1276: gap of unknown length

```

* 1277 2432: contig of 1156 bp in length
* 2433 2532: gap of unknown length
* 2533 3848: contig of 1316 bp in length
* 3849 3948: gap of unknown length
* 3949 5147: contig of 1199 bp in length
* 5148 5247: gap of unknown length
* 5248 6447: contig of 1200 bp in length
* 6448 6547: gap of unknown length
* 6548 7609: contig of 1062 bp in length
* 7610 7710: gap of unknown length
* 7710 8847: contig of 1138 bp in length
* 8848 8947: gap of unknown length
* 8948 9958: contig of 1011 bp in length
* 9959 10059: gap of unknown length
* 10059 11219: contig of 1161 bp in length
* 11220 11319: gap of unknown length
* 11320 12355: contig of 1036 bp in length
* 12356 12455: gap of unknown length
* 12456 13532: contig of 1077 bp in length
* 13533 13632: gap of unknown length
* 13633 15077: contig of 1445 bp in length
* 15078 15177: gap of unknown length
* 15178 16436: contig of 1259 bp in length
* 16437 16536: gap of unknown length
* 16537 17549: contig of 1013 bp in length
* 17550 17649: gap of unknown length
* 17650 19262: contig of 1613 bp in length
* 19263 19362: gap of unknown length
* 19363 21027: contig of 1665 bp in length
* 21028 21127: gap of unknown length
* 21128 22750: contig of 1623 bp in length
* 22751 22850: gap of unknown length
* 22851 24441: contig of 1591 bp in length
* 24442 24541: gap of unknown length
* 24542 25586: contig of 1045 bp in length
* 25587 25686: gap of unknown length
* 25687 26718: contig of 1031 bp in length
* 26718 27932: gap of unknown length
* 27933 28032: gap of unknown length
* 28033 29078: contig of 1046 bp in length
* 29079 29178: gap of unknown length
* 29179 30323: contig of 1145 bp in length
* 30324 30423: gap of unknown length
* 30424 31737: contig of 1314 bp in length
* 31738 31837: gap of unknown length
* 31838 33634: contig of 1797 bp in length
* 33635 33734: gap of unknown length
* 33735 35504: contig of 1770 bp in length
* 35505 35604: gap of unknown length
* 35605 37218: contig of 1614 bp in length
* 37219 37318: gap of unknown length
* 37319 39039: contig of 1721 bp in length
* 39040 39139: gap of unknown length
* 39140 40249: contig of 1010 bp in length
* 40250 40349: gap of unknown length
* 40350 41910: contig of 1661 bp in length
* 41911 42010: gap of unknown length
* 42011 43857: contig of 1847 bp in length
* 43858 43957: gap of unknown length
* 43959 45047: contig of 1090 bp in length
* 45048 45147: gap of unknown length
* 45148 46951: contig of 1804 bp in length
* 46952 47051: gap of unknown length
* 47052 48824: contig of 1773 bp in length
* 48825 48924: gap of unknown length
* 48925 50128: contig of 1204 bp in length
* 50129 50228: gap of unknown length
* 50229 51523: contig of 1295 bp in length
* 51524 51623: gap of unknown length
* 51624 53700: contig of 2077 bp in length
* 53701 53800: gap of unknown length
* 53801 54974: contig of 1174 bp in length

```

```

* 54975 55074: gap of unknown length
* 55075 56460: contig of 1386 bp in length
* 56461 56560: gap of unknown length
* 56561 57874: contig of 1314 bp in length
* 57875 57974: gap of unknown length
* 57975 59730: contig of 1756 bp in length
* 59731 59830: gap of unknown length
* 59831 60978: contig of 1148 bp in length
* 60979 61078: gap of unknown length
* 61079 63003: contig of 1925 bp in length
* 63004 63103: gap of unknown length
* 63104 64904: contig of 1801 bp in length
* 64905 65004: gap of unknown length
* 65005 66839: contig of 1835 bp in length
* 66840 66939: gap of unknown length
* 66940 68780: contig of 1841 bp in length
* 68781 68880: gap of unknown length
* 68881 70249: contig of 1369 bp in length
* 70250 70349: gap of unknown length
* 70350 72058: contig of 1709 bp in length
* 72059 72158: gap of unknown length
* 72159 74619: contig of 2461 bp in length
* 74620 74719: gap of unknown length
* 74720 76080: contig of 1361 bp in length
* 76081 76180: gap of unknown length
* 76181 78021: contig of 1841 bp in length
* 78022 78121: gap of unknown length
* 78122 79447: contig of 1326 bp in length
* 79448 79547: gap of unknown length
* 79548 82235: contig of 2688 bp in length
* 82236 82335: gap of unknown length
* 82336 84495: contig of 2160 bp in length

```

## Alignment Scores:

```

Pred. No.: 2.12e+03 Length: 171184
Score: 54.00 Matches: 9
Percent Similarity: 84.62% Conservative: 2
Best Local Similarity: 69.23% Mismatches: 2
Query Match: 60.00% Indels: 0
DB: 2 Gaps: 0

```

US-10-020-441-3 (1-16) x AC121203 (1-171184)

Oy 2 G1YTYG1YArgAspAspAsnAspArgAspProSerArg 14

Db 725 GCGTACGGAGACAGCATATATACACGGAAGAAAGAGAGA 687

## RESULT 10

AC107218

LOCUS Homo sapiens BAC clone RP11-561A13 from 2, complete sequence. 164620 bp DNA linear PRI 09-MAR-2002

DEFINITION AC107218

AC107218.4 GI:19071700

KEYWORDS

HTG.

SOURCE

ORGANISM

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE

1 (bases 1 to 164620)

Authors Sulston, J.E. and Waterston, R.

Title Toward a complete human genome sequence

Journal Genome Res. 8 (11), 1097-1108 (1998)

Medline 99063792

9847074  
2 (bases 1 to 164620)  
Hollmes, A. and Cotton, M.  
The sequence of Homo sapiens BAC clone RP11-561A13  
Unpublished (2001)  
3 (bases 1 to 164620)  
Waterston, R.H.  
Direct Submission  
Submitted (16-JAN-2002) Genome Sequencing Center, Washington  
University School of Medicine, 4444 Forest Park Parkway, St. Louis,

REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
MO 63108, USA  
4 (bases 1 to 164620)  
Waterston,R.H.  
Direct Submission  
Submitted (16-FEB-2002) Genome Sequencing Center, Washington  
University School of Medicine, 4444 Forest Park Parkway, St. Louis,  
MO 63108, USA  
5 (bases 1 to 164620)  
Waterston,R.H.  
Direct Submission  
Submitted (02-MAR-2002) Genome Sequencing Center, Washington  
University School of Medicine, 4444 Forest Park Parkway, St. Louis,  
MO 63108, USA  
6 (bases 1 to 164620)  
Waterston,R.  
Direct Submission  
Submitted (09-MAR-2002) Department of Genetics, Washington  
University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA  
On Mar 2, 2002 this sequence version replaced gi:18677634.  
----- Genome Center  
Center: Washington University Genome Sequencing Center  
Center code: WUGSC  
Web site: <http://genome.wustl.edu/gsc>  
Contact: [saplens@watson.wustl.edu](mailto:saplens@watson.wustl.edu)  
----- Summary Statistics  
Center project name: H\_NH0561A13  
-----

NOTICE: This sequence may not represent the entire insert of this  
clone. It may be shorter because we only sequence overlapping  
clone sections once, or longer because we provide a small overlap  
between neighboring data submissions.

This sequence was finished as follows unless otherwise noted:  
all regions were double stranded, sequenced with an alternate  
chemistry, or covered by high quality data (i.e., phred quality >=  
30); an attempt was made to resolve all sequencing problems, such  
as compressions and repeats; all regions were covered by sequence  
from more than one subclone; and the assembly was confirmed by  
restriction digest.

MAPPING INFORMATION:  
Mapping information for this clone was provided by Dr. John D.  
McPherson, Department of Genetics, Washington University, St. Louis  
MO. For additional information about the map position of this  
sequence, see <http://genome.wustl.edu/gsc>

SOURCE INFORMATION:  
The RPCI-11 human BAC library was made from the blood of one male  
donor, as described by Osoegawa,K., Moon,P.Y., Zhao,B., Frenken,E.,  
Tateno,M., Catanese,J.J. and de Jong,P.J. (1998) An improved  
approach for construction of bacterial artificial chromosome  
libraries. Genomics 51:1-8. The clone may be obtained either from  
Research Genetics, Inc. (<http://www.resgen.com>) or Pieter de Jong  
and coworkers at <http://www.chori.org>  
VECTOR: pBACe3.6

NEIGHBORING SEQUENCE INFORMATION:  
The clone sequenced to the right is RP11-172J1, 2000 bp overlap.  
Actual start of this clone is at base position 1 of RP11-561A13.

Single plasmid regions exist between 126567 and 126760; 141430 and  
141464; 105109 and 105130. Polymorphisms exist between AC06051 and  
AC107218.

FEATURES  
source  
Location/Qualifiers  
1..164620  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="2"  
/map="2"  
/clone="RP11-561A13"  
/clone\_id="RPCI-11"  
108..227  
repeat\_region

repeat\_region  
311..363  
/rpt\_family="MIR"  
repeat\_region  
504..590  
/rpt\_family="AT\_rich"  
repeat\_region  
744..767  
/rpt\_family="MIR"  
repeat\_region  
2022..2043  
/rpt\_family="AT\_rich"  
repeat\_region  
3441..3490  
/rpt\_family="AT\_rich"  
repeat\_region  
4577..4632  
/rpt\_family="AT\_rich"  
repeat\_region  
7837..7933  
/rpt\_family="AT\_rich"  
repeat\_region  
8007..8139  
/rpt\_family="MIR"  
repeat\_region  
9336..9496  
/rpt\_family="MIR"  
repeat\_region  
9568..9794  
/rpt\_family="L2"  
repeat\_region  
9817..9885  
/rpt\_family="MIR"  
repeat\_region  
9894..10015  
/rpt\_family="(TATTA)n"  
repeat\_region  
11201..11421  
/rpt\_family="L2"  
repeat\_region  
12445..12467  
/rpt\_family="Alu"  
repeat\_region  
13230..13531  
/rpt\_family="AT\_rich"  
repeat\_region  
13644..13666  
/rpt\_family="Alu"  
repeat\_region  
14616..14713  
/rpt\_family="(TG)n"  
repeat\_region  
15113..15134  
/rpt\_family="MIR"  
repeat\_region  
15375..15535  
/rpt\_family="AT\_rich"  
repeat\_region  
15615..15693  
/rpt\_family="MER1\_type"  
repeat\_region  
15743..15884  
/rpt\_family="MER103"  
repeat\_region  
16314..16514  
/rpt\_family="MIR"  
repeat\_region  
17237..17540  
/rpt\_family="MER1\_type"  
repeat\_region  
17875..18059  
/rpt\_family="Alu"  
repeat\_region  
18966..19330  
/rpt\_family="MIR"  
repeat\_region  
19355..19701  
/rpt\_family="MER2\_type"  
repeat\_region  
20159..20178  
/rpt\_family="MIR"  
repeat\_region  
21158..21467  
/rpt\_family="(A)n"  
repeat\_region  
23405..23426  
/rpt\_family="Alu"  
repeat\_region  
23541..23825  
/rpt\_family="AT\_rich"  
repeat\_region  
24084..24385  
/rpt\_family="L2"  
repeat\_region  
24504..24547  
/rpt\_family="Alu"  
repeat\_region  
25877..26058  
/rpt\_family="(TA)n"  
repeat\_region  
28533..28568  
/rpt\_family="MIR"  
repeat\_region  
29113..29153  
/rpt\_family="AT\_rich"  
repeat\_region  
29452..29672  
/rpt\_family="AT\_rich"  
repeat\_region  
/rpt\_family="CRI"

```
repeat_region      31245..31268
                    /rpt_family="AT_rich"
repeat_region      32624..32821
                    /rpt_family="L1"
repeat_region      33187..33254
                    /rpt_family="AT_rich"
repeat_region      33261..33551
                    /rpt_family="Alu"
repeat_region      33966..34008
                    /rpt_family="AT_rich"
repeat_region      34009..34201
                    /rpt_family="L1"
repeat_region      34236..34442
                    /rpt_family="MER2_type"
repeat_region      34443..34745
                    /rpt_family="Alu"
repeat_region      34746..34763
                    /rpt_family="MER2_type"
repeat_region      34789..35369
                    /rpt_family="L1"
repeat_region      35516..35609
                    /rpt_family="L1"
repeat_region      35616..35712
                    /rpt_family="L1"
repeat_region      36183..36211
                    /rpt_family="AT_rich"
repeat_region      36434..36774
                    /rpt_family="L1"
repeat_region      36829..36886
                    /rpt_family="(TA)n"
```

Alignment Scores:  
Pred. No.: 2.92e+03 Length: 164620  
Score: 53.00 Matches: 8  
Percent Similarity: 92.31% Conservative: 4  
Best Local Similarity: 61.54% Mismatches: 1  
Query Match: 58.89% Indels: 0  
DB: 9 Gaps: 0

US-10-020-441-3 (1-16) x AC107218 (1-164620)

Oy 4 G1yA-gaspaspapnasp-ArgaspProSerArglyAsn 16  
Db 41490 GGGAAAGATGACATGACACGAGAACCCAGAGAGAGTCT 41528

RESULT 11  
AB083612/c 921 bp DNA linear PRI 24-MAY-2002  
LOCUS Homo sapiens GPCR gene for putative G-protein coupled receptor,  
DEFINITION complete CDS, clone:hGPCR30.  
ACCESSION AB083612  
VERSION AB083612.1 GI:20152287  
KEYWORDS  
SOURCE Homo sapiens DNA, clone:hGPCR30.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.

REFERENCE  
AUTHORS Takeda,S., Kadowaki,S., Haga,T., Takesue,H. and Mitaku,S.  
TITLE Identification of G protein-coupled receptor genes from the human  
genome sequence  
JOURNAL Unpublished  
AUTHORS 2 (bases 1 to 921)  
TITLE Takeda,S., Kadowaki,S., Haga,T., Takesue,H. and Mitaku,S.  
JOURNAL Direct Submission  
AUTHORS Submitted (10-Apr-2002) Shigeki Takeda, Gunma University,  
TITLE Department of Biological and Chemical, Engineering, Faculty of  
JOURNAL Engineering; 1-5-1, Kiryu, Gunma 376-8515, Japan  
(E-mail:stakeda@ce.gunma-u.ac.jp, Tel:+81-277-30-1434,  
Fax:+81-277-30-1434)

FEATURES  
source 1..921  
/organism="Homo sapiens"

```
/db_xref="taxon:9606"  
/clone="hGPCR30"  
1..921  
/gene="GPCR"  
1..921  
/gene="GPCR"  
/note="predicted with SOSU analysis"
```

BASE COUNT 237 a 161 c 199 g 324 t  
ORIGIN  
Alignment Scores:  
Pred. No.: 41.8 Length: 921  
Score: 52.00 Matches: 11  
Percent Similarity: 85.71% Conservative: 1  
Best Local Similarity: 78.57% Mismatches: 1  
Query Match: 57.78% Indels: 1  
DB: 9 Gaps: 0

US-10-020-441-3 (1-16) x AB083612 (1-921)

Oy 1 ValGlyTyrGlyArgaspaspasp-ArgaspProSer 13  
Db 820 ATGGGATATGACTGACGACATGATCAGAGACCAAGT 781

RESULT 12  
AF072873/c 2697 bp mRNA linear PRI 01-JUL-1999  
LOCUS Homo sapiens frizzled 6 mRNA, complete cds.  
DEFINITION AF072873  
ACCESSION AF072873  
VERSION AF072873.1 GI:5305408  
KEYWORDS  
SOURCE Homo sapiens.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.

REFERENCE  
AUTHORS Gazit,A., Yaniv,A. and Aaronson,S.A.  
TITLE Molecular cloning of the human Frizzled 6  
JOURNAL Unpublished  
AUTHORS 2 (bases 1 to 2697)  
TITLE Gazit,A., Yaniv,A. and Aaronson,S.A.  
JOURNAL Direct Submission  
AUTHORS Submitted (18-JUN-1998) Derald H. Rittenberg Cancer Center, The  
TITLE Mount Sinai Medical Center, One Gustave L. Levy Place, East  
JOURNAL Building, New York, NY 101029-6574, USA  
location/Qualifiers

FEATURES  
source 1..2697  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
281..2401  
/function="Wnt receptor"  
/codon\_start=1  
/product="frizzled 6"  
/protein\_id="AA041637.1"  
/db\_xref="GI:5305409"  
/translation="MEMPFTLLTIFPLPLRGHSLFCEPIITPRCKMAYNMTPFPN  
LMGHYDSDIAVEMEHPLPLNLECSNIEFTLCAKAPTEIOIHVPPPCRKCEK  
YSDCKLIDTFGIMPELBCDRLQYCEIYVPVDFPTEFLGQKTEQVQDRIQGM  
CPRLKTSGGGYSFLGIDOCAPCPMMYFSDLEPAKSPIGTASIFCLCATFTFL  
TFLDVRPYPERPDIYSCVSIIVLMYIGLIDSTCKNADEKLKGLTVYVG  
SOKACTVLFMLYFTFMAGTVMYIITTFFLAAGKRSCEALQKAVHVAWGT  
PGLTVMLANLKVGNISGVCEVGLYDLSRYFVLLPLCLQVFGSLLAGIIS  
LNHVROYIOHGRNQEKIKKEMIRIGVFSGLYVLPVLLGLCYVEQVNRITWEITW

ISRLSGEOWDVGKQAGSVEKREKRSRISPKSDITDTGLAQSNLNQVSSSEPSLSLKG  
TSLSLHVPVSGVRKEQGGGSGSDT"

US-10-020-441-3 (1-16) x E31206 (1-2880)

Oy 1 ValGlyTYrGlyVArgAspAspAsnAsp-ArgAspProser 13  
Db 1614 ATGGGATATGCTAGTGCACATGATCAGAGCCCAAGT 1575

## RESULT 16

AB012911/c 3342 bp mRNA linear PRI 19-MAR-2002  
LOCUS Homo sapiens mRNA for Frizzled-6, complete cds.  
DEFINITION AB012911  
ACCESSION AB012911.1 GI:3062802  
VERSION  
KEYWORDS  
SOURCE Homo sapiens cDNA to mRNA, clone\_11b:Fetal lung cDNA library  
clone:HF6-1.

## ORGANISM

Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

## REFERENCE

1 Tokuhara,M., Hirai,M., Atomi,Y., Terada,M. and Katoh,M.  
Molecular cloning of human Frizzled-6  
JOURNAL Biochem. Biophys. Res. Commun. 243 (2), 622-627 (1998)  
MEDLINE 98153814  
REFERENCE 2 (bases 1 to 3342)  
AUTHORS Katoh,M.  
TITLE Direct Submission  
JOURNAL Submitted (10-APR-1998) Masaru Katoh, National Cancer Center  
Research Institute, Genetics Division; Tsukiji 5-chome, Chuo-ku,  
Tokyo 104-0045, Japan (E-mail:mkaoh@ncc.go.jp,  
Tel:+81-3-3542-2511(ex.4402), Fax:+81-3-3541-2685)  
Location/Qualifiers

## FEATURES

source

1..3342  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="8"  
/map="8q22.3-23.1"  
/clone="HF6-1"  
/clone\_11b="Fetal lung cDNA library"  
/note="Cloning vector:pUC118"

gene

1..3342  
/gene="Hfz6"  
259..2379  
/gene="Hfz6"  
/function="WNT receptor"  
/note="HFZ6"

CDS

/codon\_start=1  
/product="Frizzled-6"  
/protein\_id="BAA25686.1"  
/db\_xref="GI:3062803"  
/translation="MEMFTPLTLCIFPLRGHSLFCEPTITVPRCKMAVNMFPFN  
LMGHYDOSIAVEMHEPLPLANTBESPIETPLCAFPVTCIEQHIVVPRKICEKY  
YSDCKLIDTFRIRPEBELCDRLQYCDETVPVTPDPTBELGPKTBQVRDITGM  
CPRLKRTSGGGQYKFLGIDQCAPCPNNYFSDLEFAKSFIVTSIFCLCATLFTL  
TFLIDVRRFRYPERPIIYVSVCYSIVSLMFTIGFLDSTACNKADEKLEIGDITVLG  
SONKACTVLMFLYFTMAGVVMWVILITITWFLAAGKMSCEALBOKAVMFAVAMGT  
PGFLTVMLLANKKVEGNIISGVCFVGYADLASRFVLLPCLCVFVGLSLLAGIIS  
LNHYROYIHDGRNOEKLKKEMIRIGVPSGLYIVPLVITLIGCYVVEOYNRIITWETW  
SDHCQYHIPCPYQAKAKARPELALPMIKYIMTLIVGISAVFWGSKTCTEMAGFPK  
NRRKRDPISESRRVLOESCEFFLHNSKVKHKHKHYKPSHKLKVISKWTSGATA  
NHGTSVAVALTSHDYLGOETLTELITQTSPEISMREYVADGASTPRLBEDCCGPAPPAAS  
ISRLSGEVDKGGAGSVSESARSEGRISPKSDITDTGLAQSNMLQVSSSEPSLKG  
STSLVHPVSGVRKEGGGCHSDT"

BASE COUNT 981 a 621 c 688 g 1052 t  
ORIGIN

## Alignment Scores:

Pred. No.: 131 Length: 3342  
Score: 52.00 Matches: 11  
Percent Similarity: 85.71% Conservatve: 1  
Best Local Similarity: 78.57% Mismatches: 1  
Query Match: 57.78% Indels: 0  
DB: 9 Gaps: 0

US-10-020-441-3 (1-16) x AB012911 (1-3342)

Oy 1 ValGlyTYrGlyVArgAspAspAsp-ArgAspProser 13  
Db 1639 ATGGGATATGCTAGTGCACATGATCAGAGCCCAAGT 1600

## RESULT 17

AB065702/c 31808 bp DNA linear PRI 23-JUL-2002  
LOCUS Homo sapiens gene for seven transmembrane helix receptor, complete  
DEFINITION cds, isolate:CBRC7TM\_265.  
ACCESSION AB065702  
VERSION AB065702.1 GI:21928674  
KEYWORDS  
SOURCE Homo sapiens (isolate:CBRC7TM\_265) DNA.

## ORGANISM

Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

## REFERENCE

1 Suwa,M., Saco,T., Okouchi,I., Arita,M., Futami,K., Matsumoto,S.,  
Tsubsumi,S., Aburatani,H., Asai,K. and Akiyama,Y.  
Genome-wide discovery and analysis of human seven transmembrane  
helix receptor genes  
JOURNAL Unpublished  
REFERENCE 2 (bases 1 to 31808)  
AUTHORS Suwa,M.  
TITLE Direct Submission

JOURNAL Submitted (11-JUL-2001) Makiko Suwa, Computational Biology Research  
Center (CBRC), National Institute of Advanced Industrial Science  
and Technology (AIST); 2-41-6 Aomi Koto-ku, Tokyo 135-0064, Japan  
(E-mail:m-suwa@aist.go.jp, URL:http://www.cbrc.jp/,  
Tel:81-3-3599-8080, Fax:81-3-3599-8081)

## COMMENT

This sequence is a seven transmembrane helix receptor candidate  
predicted from the whole human genome sequences using our automated  
system that contains programs of gene  
finding (GeneDecoder), sequence search, motif-domain assignment and  
transmembrane helix prediction.  
And the sequence is submitted by the collaborative project between  
[Computational Biology Research Center (CBRC), National Institute  
of Advanced Industrial Science and Technology (AIST)] and [Genome  
Science Division, Research Center for Advanced Science and  
Technology (RCAST), University of Tokyo].  
Location/Qualifiers

## FEATURES

source

1..31808  
/organism="Homo sapiens"  
/isolate="CBRC7TM\_265"  
/db\_xref="taxon:9606"  
/chromosome="8"  
join(201..377,18675..18871,24568..25585,28354..28502,  
29754..30164,31440..31608)  
/codon\_start=1  
/evidence="not\_experimental"  
/product="seven transmembrane helix receptor"  
/protein\_id="BAC05925.1"  
/db\_xref="GI:21928675"

CDS

/translation="MEMFTPLTLCIFPLRGHSLFCEPTITVPRCKMAVNMFPFN  
LMGHYDOSIAVEMHEPLPLANTBESPIETPLCAFPVTCIEQHIVVPRKICEKY  
YSDCKLIDTFRIRPEBELCDRLQYCDETVPVTPDPTBELGPKTBQVRDITGM  
CPRLKRTSGGGQYKFLGIDQCAPCPNNYFSDLEFAKSFIVTSIFCLCATLFTL  
TFLIDVRRFRYPERPIIYVSVCYSIVSLMFTIGFLDSTACNKADEKLEIGDITVLG  
SONKACTVLMFLYFTMAGVVMWVILITITWFLAAGKMSCEALBOKAVMFAVAMGT  
PGFLTVMLLANKKVEGNIISGVCFVGYADLASRFVLLPCLCVFVGLSLLAGIIS  
LNHYROYIHDGRNOEKLKKEMIRIGVPSGLYIVPLVITLIGCYVVEOYNRIITWETW  
SDHCQYHIPCPYQAKAKARPELALPMIKYIMTLIVGISAVFWGSKTCTEMAGFPK  
NRRKRDPISESRRVLOESCEFFLHNSKVKHKHKHYKPSHKLKVISKWTSGATA  
NHGTSVAVALTSHDYLGOETLTELITQTSPEISMREYVADGASTPRLBEDCCGPAPPAAS  
ISRLSGEVDKGGAGSVSESARSEGRISPKSDITDTGLAQSNMLQVSSSEPSLKG  
STSLVHPVSGVRKEGGGCHSDT"

BASE COUNT 9628 a 5585 c 6164 g 10431 t  
ORIGIN

## Alignment Scores:



Pred. No.: 968 Length: 31808  
 Score: 52.00 Matches: 11  
 Percent Similarity: 85.71% Conservative: 1  
 Best Local Similarity: 78.57% Mismatches: 1  
 Query Match: 57.78% Indels: 1  
 DB: 9 Gaps: 0

US-10-020-441-3 (1-16) x AB065702 (1-31808)

Cy 1 ValGlyTYrGTAAGAspAspAspAsp-ArgAspProSer 13

Dn 25574 ATGGGATATGGTACTGACGACATATGACAGACCCAACT 25535

RESULT 18

AC069351/c

LOCUS AC069351 111496 bp DNA linear PRI 17-FEB-2002  
 DEFINITION Homo sapiens chromosome 8, clone RP11-432K23, complete sequence.  
 ACCESSION AC069351  
 VERSION AC069351.5 GI:18698795  
 KEYWORDS HTG.

SOURCE

ORGANISM

Homo sapiens.  
 Homo sapiens.  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE  
 1 (bases 1 to 111496)  
 Birren, B., Linton, L., Nusbaum, C. and Lander, E.

2 (bases 1 to 111496)  
 Birren, B., Linton, L., Nusbaum, C., Lander, E., Abrahams, H., Allen, N.,

Anderson, S., Baldwin, J., Barna, N., Baetien, V., Beda, F.,  
 Boguslavskiy, L., Bouckhgalter, B., Brown, A., Burkett, G.,

Campioano, A., Cooke, P., Dearrellano, K., Dewar, K., Diaz, J.S.,  
 Dodge, S., Donno, M., Doyle, M., Ferreira, P., Fitzhugh, W., Gage, D.,

Galagan, J., Gardyna, S., Ginde, S., Goyette, M., Graham, L.,  
 Grand-Pierre, N., Grant, G., Hagos, B., Heaford, A., Horton, L.,

Holland, J.C., Iliev, I., Johnson, R., Jones, C., Kam, L., Karatas, A.,  
 Klein, J., Laroque, K., Lamazares, R., Landers, T., Lehoczy, J.,

Levine, R., Liu, C., Liu, G., Locke, K., MacDonald, P., Margulis, N.,  
 McCarthy, M., McEwan, P., McGurk, A., McKernan, K., McPheters, R.,

Meldrum, J., Menus, L., Minova, T., Miranda, C., Mlenga, V., Morrow, J.,  
 Murphy, T., Naylor, J., Norman, C.H., O'Connor, T., O'Donnell, P.,

O'Neil, D., Oliver, T.M., Oliver, J., Peterson, K., Pierre, N.,  
 Pisan, C., Pollara, V., Raymond, C., Riley, R., Rogov, P., Rothman, D.,

Roy, A., Santos, R., Schauer, S., Severy, P., Spencer, B.,  
 Stange-Thomann, N., Stojanovic, N., Subramanian, A., Talamas, J.,

Tesfaye, S., Theodore, J., Tirrell, A., Travers, M., Triggillo, J.,  
 Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W.J.,

Young, G., Zainoun, J., Zimmer, A. and Zody, M.

Direct Submission

Submitted (30-MAY-2000) Whitehead Institute/MIT Center for Genome  
 Research, 320 Charles Street, Cambridge, MA 02141, USA

3 (bases 1 to 111496)

Anderson, S., Barna, N., Nusbaum, C., Lander, E., Ali, A., Allen, N.,  
 Birren, B., Linton, L., Boguslavskiy, L., Bouckhgalter, B.,

Brown, A., Camarata, J., Campioano, A., Chang, J., Chazaro, B.,  
 Choepel, Y., Colangelo, M., Collins, S., Collymore, A., Cook, A.,

Cooke, P., Dearrellano, K., Dewar, K., Diaz, J.S., Dodge, S., Fato, S.,  
 Ferreira, P., Fitzhugh, W., Gage, D., Galagan, J., Gardyna, S.,

Ginde, S., Gord, S., Goyette, M., Graham, L., Grand-Pierre, N.,  
 Hagos, B., Heaford, A., Horton, L., Hulme, W., Iliev, I., Johnson, R.,

Jones, C., Kamat, A., Karatas, A., Kells, C., Laroque, K.,  
 Lamazares, R., Landers, T., Lehoczy, J., Levine, R., Liu, G.,

Maclean, C., MacDonald, P., Major, J., Marguis, N., Matthews, C.,  
 McCarthy, M., McEwan, P., McKernan, K., McPheters, R., Meldrum, J.,

Menus, L., Minova, T., Mlenga, V., Murphy, T., Naylor, J., Nguyen, C.,  
 Norbu, C., Norman, C.H., O'Connor, T., O'Donnell, P., O'Neil, D.,

Oliver, J., Peterson, K., Phunhang, P., Pierre, N., Pollara, V.,  
 Raymond, C., Retta, R., Rieback, M., Riley, R., Rise, C., Rogov, P.,

Roman, J., Rosetti, M., Roy, A., Santos, R., Schauer, S., Schupack, R.,  
 Seaman, S., Severy, P., Spencer, B., Stange-Thomann, N., Stojanovic, N.,

Straus, N., Subramanian, A., Talamas, J., Tesfaye, S., Theodore, J.,

Topham, K., Travers, M., Travis, N., Triggillo, J., Vassiliev, H.,  
 Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W.J., Young, G.,  
 Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.

Direct Submission

Submitted (01-SEP-2001) Whitehead Institute/MIT Center for Genome  
 Research, 320 Charles Street, Cambridge, MA 02141, USA

4 (bases 1 to 111496)

Anderson, S., Barna, N., Nusbaum, C., Lander, E., Ali, A., Allen, N.,  
 Birren, B., Linton, L., Boguslavskiy, L., Bouckhgalter, B.,

Brown, A., Camarata, J., Campioano, A., Chang, J., Chazaro, B.,  
 Choepel, Y., Colangelo, M., Collins, S., Collymore, A., Cook, A.,

Cooke, P., Dearrellano, K., Dewar, K., Diaz, J.S., Dodge, S., Fato, S.,  
 Ferreira, P., Fitzhugh, W., Gage, D., Galagan, J., Gardyna, S.,

Ginde, S., Gord, S., Goyette, M., Graham, L., Grand-Pierre, N.,  
 Hagos, B., Heaford, A., Horton, L., Hulme, W., Iliev, I., Johnson, R.,

Jones, C., Kamat, A., Karatas, A., Kells, C., Laroque, K.,  
 Lamazares, R., Landers, T., Lehoczy, J., Levine, R., Liu, G.,

Maclean, C., MacDonald, P., Major, J., Marguis, N., Matthews, C.,  
 McCarthy, M., McEwan, P., McKernan, K., McPheters, R., Meldrum, J.,

Menus, L., Minova, T., Mlenga, V., Murphy, T., Naylor, J., Nguyen, C.,  
 Norbu, C., Norman, C.H., O'Connor, T., O'Donnell, P., O'Neil, D.,

Oliver, J., Peterson, K., Phunhang, P., Pierre, N., Pollara, V.,  
 Raymond, C., Retta, R., Rieback, M., Riley, R., Rise, C., Rogov, P.,

Roman, J., Rosetti, M., Roy, A., Santos, R., Schauer, S., Schupack, R.,  
 Seaman, S., Severy, P., Spencer, B., Stange-Thomann, N., Stojanovic, N.,

Straus, N., Subramanian, A., Talamas, J., Tesfaye, S., Theodore, J.,  
 Topham, K., Travers, M., Travis, N., Triggillo, J., Vassiliev, H.,

Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W.J., Young, G.,  
 Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.

Direct Submission

Submitted (03-SEP-2001) Whitehead Institute/MIT Center for Genome  
 Research, 320 Charles Street, Cambridge, MA 02141, USA

5 (bases 1 to 111496)

Birren, B., Linton, L., Nusbaum, C., Lander, E., Ali, A., Allen, N.,  
 Anderson, S., Barna, N., Baetien, V., Boguslavskiy, L., Bouckhgalter, B.,

Brown, A., Camarata, J., Campioano, A., Chang, J., Chazaro, B.,  
 Choepel, Y., Colangelo, M., Collins, S., Collymore, A., Cook, A.,

Cooke, P., Dearrellano, K., Dewar, K., Diaz, J.S., Dodge, S., Fato, S.,  
 Ferreira, P., Fitzhugh, W., Gage, D., Galagan, J., Gardyna, S.,

Ginde, S., Gord, S., Goyette, M., Graham, L., Grand-Pierre, N.,  
 Hagos, B., Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C.,

Kamat, A., Karatas, A., Kells, C., Laroque, K., Lamazares, R.,  
 Landers, T., Lehoczy, J., Levine, R., Liu, G., Maclean, C.,

Macdonald, P., Major, J., Marguis, N., Matthews, C., McCarthy, M.,  
 McEwan, P., McKernan, K., Meldrum, J., Menus, L., Minova, T.,

Mlenga, V., Murphy, T., Naylor, J., Nguyen, C., Nicol, R., Norbu, C.,  
 Norman, C.H., O'Connor, T., O'Donnell, P., O'Neil, D., Oliver, J.,

Peterson, K., Phunhang, P., Pierre, N., Pollara, V., Raymond, C.,  
 Retta, R., Rieback, M., Riley, R., Rise, C., Rogov, P., Roman, J.,

Rosetti, M., Roy, A., Santos, R., Schauer, S., Schupack, R., Seaman, S.,  
 Severy, P., Spencer, B., Stange-Thomann, N., Stojanovic, N.,

Straus, N., Subramanian, A., Talamas, J., Tesfaye, S., Theodore, J.,  
 Topham, K., Travers, M., Travis, N., Triggillo, J., Vassiliev, H.,

Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W.J., Young, G.,  
 Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.

Direct Submission

Submitted (17-FEB-2002) Whitehead Institute/MIT Center for Genome  
 Research, 320 Charles Street, Cambridge, MA 02141, USA

On Feb 17, 2002 this sequence version replaced ref:15412469.  
 All repeats were identified using RepeatMasker:  
 Smt, A.F.A. & Green, P. (1996-1997)  
 http://ftp.genome.washington.edu/RM/RepeatMasker.html

Genome Center  
 Center: Whitehead Institute/ MIT Center for Genome Research  
 Center code: WIGBR  
 Web site: http://www-seq.wi.mit.edu  
 Contact: sequence\_submissions@genome.wi.mit.edu  
 Project Information  
 Center project name: LI0384  
 Center clone name: 432\_K\_23  
 Location/Qualifiers  
 1..111496

FEATURES

source

```

/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="8"
/map="8"
/clone="RP11-432K23"
/clone_lib="RP11-11 Human Male BAC"
2. .1076
repeat_region /rpt_family="MSTA-int"
1078. .1432
repeat_region /rpt_family="MSTA"
1422. .1484
repeat_region /rpt_family="THE1B"
complement(1490. .2331)
repeat_region /rpt_family="L1M4"
complement(12333. .2799)
repeat_region /rpt_family="L1M4"
complement(12828. .3015)
repeat_region /rpt_family="AluDo"
complement(3017. .3934)
repeat_region /rpt_family="L1M4"
complement(4459. .4932)
repeat_region /rpt_family="L1M4"
complement(5045. .5413)
repeat_region /rpt_family="L1M4"
5869. .5891
repeat_region /rpt_family="AT rich"
complement(6102. .6275)
repeat_region /rpt_family="L1R3A"
6587. .6765
repeat_region /rpt_family="MER3"
complement(6726. .6771)
repeat_region /rpt_family="MER3"
6852. .7062
repeat_region /rpt_family="MIR"
complement(8954. .9162)
repeat_region /rpt_family="MIR"
9211. .9345
repeat_region /rpt_family="AluSq/x"
9346. .9640
repeat_region /rpt_family="AluSx"
complement(9744. .9940)
repeat_region /rpt_family="MIR"
10139. .10194
repeat_region /rpt_family="L3"
10333. .10432
repeat_region /rpt_family="L3"
10433. .10858
repeat_region /rpt_family="MT1H"
complement(11026. .11269)
repeat_region /rpt_family="MIR"
12494. .12526
repeat_region /rpt_family="(TCTC)n"
complement(12532. .12887)
repeat_region /rpt_family="L1MC1"
complement(12888. .13199)
repeat_region /rpt_family="AluSp"
complement(13200. .13754)
repeat_region /rpt_family="L1MC1"
14817. .14849
repeat_region /rpt_family="(TAA)n"
complement(15078. .15216)
repeat_region /rpt_family="MER69A"
complement(15499. .15654)
repeat_region /rpt_family="L1MA9"
15655. .15703
repeat_region /rpt_family="(TG)n"
complement(15704. .15809)
repeat_region /rpt_family="L1MA9"
complement(15810. .16107)
repeat_region

Alignment Scores: 2.95e+03 Length: 111496
Pred. No.: 52.00 Matches: 11

```

```

Percent Similarity: 85.71% Conservative: 1
Best Local Similarity: 78.57% Mismatches: 1
Query Match: 57.78% Indels: 1
DB: 9 Gaps: 0

US-10-020-441-3 (1-16) x AC0069351 (1-111496)
Qy 1 ValGlyTyrGlyArgAspAspAspArgAspProSer 13
Db 54401 ATGGGATATGTGTTACTGACGACATGATCATGACCAAGT 54362

RESULT 19
AP004221/c 160197 bp DNA linear PRI 02-OCT-2001
LOCUS Homo sapiens genomic DNA, chromosome 8q23, clone: KB1896D3.
DEFINITION AP004221
ACCESSION AP004221
VERSION AP004221.1 GI:15824052
KEYWORDS
SOURCE
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE
1 Shimizu,N. and Asakawa,S.
Homo sapiens DNA chromosome 8 SEQUENCE
JOURNAL Published Only in Database (2001)
2 (bases 1 to 160197)
AUTHORS Shimizu,N. and Asakawa,S.
TITLE Direct Submission
JOURNAL Submitted (28-SEP-2001) Nobuyoshi Shimizu, Keio University, School
of Medicine, Molecular Biology/ 35 Shinanomachi, Shinjuku-ku, Tokyo
160-8582, Japan (E-mail:nshimizu@med.keio.ac.jp,
Tel:81-3-3351-2370, Fax:81-3-3351-2370)

FEATURES
source
1..160197
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="8"
/map="8q23"
/clone="KB1896D3"
/cell_line="FHEB 14 - 14"
/clone_lib="Keio BAC library"
1. .601
repeat_region /evidence=not experimental
/rpt_family="L1M4"
602. .893
repeat_region /evidence=not experimental
/rpt_family="AluDo"
896. .1210
repeat_region /evidence=not experimental
/rpt_family="AluSp"
1211. .1569
repeat_region /evidence=not experimental
/rpt_family="L1M4"
1570. .1877
repeat_region /evidence=not experimental
/rpt_family="AluVag8"
1878. .2851
repeat_region /evidence=not experimental
/rpt_family="L1M4"
2852. .3150
repeat_region /evidence=not experimental
/rpt_family="AluDo"
3151. .3431
repeat_region /evidence=not experimental
/rpt_family="L1M4"
3479. .3568
repeat_region /evidence=not experimental
/rpt_family="FLM_A"
3594. .4263
repeat_region /evidence=not experimental
/rpt_family="L1M4"

```

repeat_region	4260..4424	/evidence=not experimental	repeat_region	/evidence=not experimental
repeat_region	/rpt_family="LjMB7"	complement(4425..4719)	repeat_region	/rpt_family="Alusq/x"
repeat_region	/evidence=not experimental	/rpt_family="Alusx"	repeat_region	/evidence=not experimental
repeat_region	4720..4865	/evidence=not experimental	repeat_region	/rpt_family="AlusC"
repeat_region	/rpt_family="LjMB7"	complement(6296..6499)	repeat_region	/evidence=not experimental
repeat_region	/evidence=not experimental	/rpt_family="L2"	repeat_region	/rpt_family="MLT1H"
repeat_region	complement(7015..7210)	/evidence=not experimental	repeat_region	complement(23050..23293)
repeat_region	/rpt_family="MER5A"	complement(7558..7769)	repeat_region	/evidence=not experimental
repeat_region	/evidence=not experimental	/rpt_family="LTR16C"	repeat_region	/rpt_family="MIR"
repeat_region	9650..9870	/evidence=not experimental	repeat_region	24518..24550
repeat_region	/rpt_family="MIR"	complement(110491..11094)	repeat_region	/evidence=not experimental
repeat_region	/evidence=not experimental	/rpt_family="LjMB8"	repeat_region	/rpt_family="CTC/n"
repeat_region	11110..11481	/evidence=not experimental	repeat_region	/evidence=not experimental
repeat_region	/rpt_family="MSTA"	complement(114851..15038)	repeat_region	/rpt_family="LjMC1"
repeat_region	11482..13100	/evidence=not experimental	repeat_region	26835..26871
repeat_region	/rpt_family="MSTA-internal"	complement(13514..14354)	repeat_region	/evidence=not experimental
repeat_region	13102..13456	/evidence=not experimental	repeat_region	/rpt_family="TTAA/n"
repeat_region	/rpt_family="LjMB8"	complement(14356..14822)	repeat_region	complement(27100..27238)
repeat_region	/evidence=not experimental	/rpt_family="LjMB4"	repeat_region	/evidence=not experimental
repeat_region	complement(14851..15038)	/evidence=not experimental	repeat_region	/rpt_family="MER9A"
repeat_region	/rpt_family="Alufo"	complement(15040..15957)	repeat_region	complement(27521..27676)
repeat_region	/evidence=not experimental	/rpt_family="LjMB4"	repeat_region	/evidence=not experimental
repeat_region	complement(16482..16711)	/evidence=not experimental	repeat_region	/rpt_family="LjMB9"
repeat_region	/rpt_family="LjMB4"	complement(16776..16955)	repeat_region	complement(27726..27831)
repeat_region	/evidence=not experimental	/rpt_family="LjMB4"	repeat_region	/evidence=not experimental
repeat_region	complement(17068..17436)	/evidence=not experimental	repeat_region	/rpt_family="LjMB9"
repeat_region	/rpt_family="LjMB4"	complement(17927..17992)	repeat_region	complement(28130..28613)
repeat_region	/evidence=not experimental	/rpt_family="T-rich"	repeat_region	/evidence=not experimental
repeat_region	complement(18186..18288)	/evidence=not experimental	repeat_region	/rpt_family="LjMB9"
repeat_region	/rpt_family="LTR50"	complement(18610..18788)	repeat_region	complement(28621..29393)
repeat_region	/evidence=not experimental	/rpt_family="MER3"	repeat_region	/evidence=not experimental
repeat_region	18875..19085	/evidence=not experimental	repeat_region	/rpt_family="Tiger1"
repeat_region	/rpt_family="MIR"	complement(20978..21186)	repeat_region	complement(31283..32038)
repeat_region	/evidence=not experimental	/rpt_family="MIR"	repeat_region	/evidence=not experimental
repeat_region	21235..21369	/evidence=not experimental	repeat_region	complement(32167..32476)
			repeat_region	/rpt_family="LjMB9"
			repeat_region	32511..34245
			repeat_region	/evidence=not experimental

  

Alignment Scores:	4.07e+03	Length:	160197
Pred. No.:	52.00	Matches:	11
Percent Similarity:	85.71%	Conservative:	1
Best Local Similarity:	78.57%	Mismatches:	1

Query Match:	57.78%	Indels:	1
DB:	9	Gaps:	0
US-10-020-441-3 (1-16) x AP004221 (1-160197)			
<p>1 ValG1YTYRGLYARGASPAASPAASAP-ARGASPPROSER 13</p> <p>Db 66418 ATGGATATGCTACTGACCAATGATGATCAGAGCCCAAGT 66379</p>			
RESULT 20			
LOCUS	AC025370	193589 bp	DNA linear PRI 20-FEB-2002
DEFINITION	Homo sapiens chromosome 8, clone RP11-318M2, complete sequence.		
ACCESSION	AC025370	GI:18767552	
VERSION	AC025370.12		
KEYWORDS	HTG.		
SOURCE	Homo sapiens.		
ORGANISM	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.		
REFERENCE	1 (bases 1 to 193589)		
AUTHORS	Birren, B., Linton, L., Nusbaum, C. and Lander, E.		
TITLE	Unpublished		
JOURNAL	2 (bases 1 to 193589)		
REFERENCE	Birren, B., Linton, L., Nusbaum, C., Lander, E., Abraham, H., Allen, N., Anderson, S., Baldwin, J., Barna, N., Bastien, V., Beda, F., Boguslavsky, L., Borkhagalter, B., Brown, A., Burkett, G., Campopiano, A., Castle, A., Choepel, Y., Colangelo, M., Collins, S., Collamore, A., Cooke, P., DeArrellano, K., Dewar, K., Diaz, J.S., Dodge, S., Domino, M., Doyle, M., Ferreira, P., Fitzhugh, W., Gage, D., Galagan, J., Gardyna, S., Ginde, S., Goyette, M., Graham, L., Grand-Pierre, N., Grant, G., Hagos, B., Heatford, A., Horton, L., Howland, J.C., Iliev, I., Johnson, R., Jones, C., Kann, L., Karatas, A., Klein, J., Lacroque, K., Lamazares, R., Landers, T., Lehoczy, J., Levine, R., Liu, C., Liu, G., Locke, K., Macdonald, P., Margitis, N., McCarthy, M., McEwan, P., McGurk, A., McKernan, K., McNetters, R., Meldrum, J., Meneus, L., Mihova, T., Miranda, C., Mlenga, V., Morrow, J., Murphy, T., Naylor, J., Norman, C.H., O'Connor, T., O'Donnell, P., O'Neill, D., Ollivar, T.M., Oliver, J., Peterson, K., Pierre, N., Pisan, C., Pollara, V., Raymond, C., Riley, R., Rogov, P., Rothman, D., Roy, A., Santos, R., Schauer, S., Severy, P., Spencer, B., Stange-Thomann, N., Scjlanovic, N., Subramanian, A., Talamas, J., Testa, S., Theodore, J., Tirrell, A., Travers, M., Trigilio, J., Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W. J., Young, G., Zainoun, J., Zimmer, A. and Zody, M.		
TITLE	Direct Submission		
JOURNAL	Submitted (08-MAR-2000) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA		
REFERENCE	3 (bases 1 to 193589)		
AUTHORS	Birren, B., Linton, L., Nusbaum, C., Lander, E., Ali, A., Allen, N., Anderson, S., Barna, N., Bastien, V., Boguslavsky, L., Borkhagalter, B., Brown, A., Camarata, J., Campopiano, A., Chang, J., Chazaro, B., Choepel, Y., Colangelo, M., Collins, S., Collamore, A., Cook, A., Cooke, P., DeArrellano, K., Dewar, K., Diaz, J.S., Dodge, S., Faro, S., Ferreira, P., Fitzhugh, W., Gage, D., Galagan, J., Gardyna, S., Ginde, S., Gort, S., Goyette, M., Graham, L., Grand-Pierre, N., Hagos, B., Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C., Kamat, A., Karatas, A., Kells, C., Lacroque, K., Lamazares, R., Landers, T., Lehoczy, J., Levine, R., Liu, G., Maclean, C., Macdonald, P., Major, J., Margus, N., Matthews, C., McCarthy, M., McEwan, P., McKernan, K., Meldrum, J., Meneus, L., Mihova, T., Mlenga, V., Murphy, T., Naylor, J., Nguyen, C., Nicol, R., Norbu, C., Norman, C.H., O'Connor, T., O'Donnell, P., O'Neill, D., Oliver, J., Peterson, K., Plunkhang, P., Pierre, N., Pollara, V., Raymond, C., Retta, R., Rieback, M., Riley, R., Rise, C., Rogov, P., Roman, J., Rosetti, M., Roy, A., Santos, R., Schauer, S., Schnapack, R., Seaman, S., Severy, P., Spencer, B., Stange-Thomann, N., Scjlanovic, N., Strauss, N., Subramanian, A., Talamas, J., Testa, S., Theodore, J., Topham, K., Travers, M., Travis, N., Trigilio, J., Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W. J., Young, G., Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.		
TITLE	Direct Submission		

JOURNAL Submitted (15-FEB-2002) Whitehead Institute/MIT Center for Genomic Research, 320 Charles Street, Cambridge, MA 02141, USA

REFERENCE 4 (bases 1 to 193589)

AUTHORS Birren, B., Linton, L., Nusbaum, C., Lander, E., Ali, A., Allen, N., Anderson, S., Barta, N., Bastien, V., Boguslavsky, L., Boukhgalter, B., Brown, A., Camarata, J., Campione, A., Chang, J., Chazaro, B., Choquet, Y., Colangelo, M., Collins, S., Collumore, A., Cook, A., Cooke, P., D'Arrelano, K., Dewar, K., Diaz, J. S., Dodge, S., Faro, S., Ferreira, P., Fitzhugh, W., Gage, D., Galagan, D., Gaidyna, S., Ginde, S., Gord, S., Guyette, M., Graham, L., Grand-Pierre, N., Hagos, B., Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C., Kamat, A., Karatas, A., Kells, C., LaRocque, K., Lamazares, R., Landers, T., Lehoccky, J., Levine, R., Liu, G., Maclean, C., Macdonald, P., Major, J., Margis, N., Matthews, C., McCarthy, M., McKean, P., McKernan, K., Melitini, J., Meneus, L., Mihova, T., Mlenga, V., Murphy, T., Naylor, J., Nguyen, C., Nicol, R., Notbu, C., Norman, C. H., O'Connor, T., O'Donnell, P., O'Neill, D., Oliver, J., Peterson, K., Plunkhang, P., Pierre, N., Pollara, V., Raymond, C., Retta, R., Riebeck, M., Riley, R., Rise, C., Rogov, P., Roman, J., Rosetti, M., Roy, A., Santos, R., Schauer, S., Schnupack, R., Seaman, S., Severy, P., Spencer, B., Strange-Thomann, N., Stojanovic, N., Strauss, N., Subramanian, A., Talamas, J., Testaye, S., Theodore, J., Tophan, K., Travers, M., Travis, N., Trifilio, D., Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W. J., Young, G., Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.

TITLE Direct Submission

JOURNAL Submitted (20-FEB-2002) Whitehead Institute/MIT Center for Genomic Research, 320 Charles Street, Cambridge, MA 02141, USA

COMMENT On Feb 20, 2002 this sequence version replaced gi:18677536. All repeats were identified using RepeatMasker: Smit, A.F.A. & Green, P. (1996-1997) <http://ftp.genome.washington.edu/RM/RepeatMasker.html>

----- Genome Center

Center: Whitehead Institute/ MIT Center for Genome Research

Center code: MIBR

Web site: <http://www-seq.wi.mit.edu>

----- Project Information

Center project name: L7927

Center clone name: 318\_M\_2

-----

FEATURES

source

-----

1.193589

/organism="Homo sapiens"

/db\_xref="taxon:9606"

/chromosome="8"

/map="8"

/clone="RP11-318M2"

/clone\_11b="RP11 Human Male BAC"

310..344

/rpt\_family="AT\_rich"

complement(1581..1794)

/rpt\_family="A1USg/x"

1805..1827

/rpt\_family="(T)n"

1866..1886

/rpt\_family="AT\_rich"

1994..2103

/rpt\_family="L1M4"

complement(2259..2561)

/rpt\_family="A1uv"

2791..3012

/rpt\_family="L1M4"

3218..3462

/rpt\_family="HAL1"

complement(3488..3944)

/rpt\_family="L1MC3"

complement(6586..6878)

/rpt\_family="A1USC"

complement(7001..7223)

/rpt\_family="L2"

8742..8763

/rpt\_family="AT\_rich"

```
repeat_region      9917..9940
                    /rpt_family="AT_rich"
repeat_region      complement(10150..10461)
                    /rpt_family="AluSx"
repeat_region      complement(10835..10994)
                    /rpt_family="L1MC/D"
repeat_region      11686..11728
                    /rpt_family="AT_rich"
repeat_region      14044..14340
                    /rpt_family="AluSx"
repeat_region      14537..14625
                    /rpt_family="AT_rich"
repeat_region      complement(14650..14924)
                    /rpt_family="L1ME3"
repeat_region      complement(15385..15507)
                    /rpt_family="FLAM_A"
repeat_region      16721..16897
                    /rpt_family="L2"
repeat_region      16924..17763
                    /rpt_family="L2"
repeat_region      18396..18900
                    /rpt_family="L2"
repeat_region      19217..19251
                    /rpt_family="AT_rich"
repeat_region      19386..19920
                    /rpt_family="L2"
repeat_region      complement(19921..20149)
                    /rpt_family="AluS5/x"
repeat_region      20150..20920
                    /rpt_family="L2"
repeat_region      20929..21976
                    /rpt_family="HL1"
repeat_region      complement(21996..22415)
                    /rpt_family="L1ME1"
repeat_region      complement(22416..22605)
                    /rpt_family="L2"
repeat_region      22779..22872
                    /rpt_family="L2"
repeat_region      complement(23728..23760)
                    /rpt_family="Charlie8"
repeat_region      complement(23795..23929)
                    /rpt_family="Charlie8"
repeat_region      23971..24013
                    /rpt_family="(CA)n"
repeat_region      complement(24078..24117)
                    /rpt_family="L3"
repeat_region      complement(24527..24745)
                    /rpt_family="MIR"
repeat_region      25840..25936
                    /rpt_family="MIR3"
repeat_region      complement(26072..26207)
                    /rpt_family="FRAM"
repeat_region      complement(26220..26316)
                    /rpt_family="AluJ/FLAM"
repeat_region      26694..26807
                    /rpt_family="LTR67"
repeat_region      27210..27495
                    /rpt_family="AluO"
repeat_region      28330..28522
                    /rpt_family="HERVL"
repeat_region      28523..28703
                    /rpt_family="MER5A"
repeat_region      29279..29861
                    /rpt_family="HERVL"
repeat_region      29930..30020
```

```
Alignment Scores:
Pred. No.: 4.81e+03      Length: 193589
Score: 52.00           Matches: 11
Percent Similarity: 85.71%      Conservative: 1
Best Local Similarity: 78.57%    Mismatches: 1
Query Match: 57.78%           Indels: 1
DB: 9                   Gaps: 0
```

```
US-10-020-441-3 (1-16) x AC025370 (1-193589)
Cy      1 ValGlyTyrGlyArgAspAspAsp-ArgAspProSer 13
Db 12287 ATGGGATATGTAAGTACGACATGATCAAGACCCAAAGT 12326

RESULT 21
AF363578/c
LOCUS
DEFINITION
AF363578      347253 bp      DNA      linear      PRI 27-DEC-2001
Homo sapiens ATPase H+ transporting lysosomal protein (ATP6C) gene,
complete cds; BAluC (BAluC) gene, complete cds, alternatively
spliced; and frizzled-like protein 6 (FZD6) gene, complete cds.
AF363578
AF363578.1 GI:17979892
SOURCE
ORGANISM
Homo sapiens.
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE
1 (bases 1 to 347253)
Tanner,S.M., Austin,J.L., Leone,G., Rush,L.J., Plasse,C.,
Heinonen,K., Wrozek,K., Sill,H., Knuutila,S., Kolitz,J.E.,
Archer,K.J., Calligaris,M.A., Bloomfield,C.D. and de la Chapelle,A.
BAluC, the human member of a novel mammalian neuroectoderm gene
lineage, is implicated in hematopoiesis and acute leukemia
Proc. Natl. Acad. Sci. U.S.A. 98 (24), 13901-13906 (2001)
JOURNAL
MEDLINE
21574584
PUBMED
11707601
REFERENCE
Tanner,S.M., Austin,J.L. and de la Chapelle,A.
AUTHORS
Direct Submission
JOURNML
Submitted (22-MAR-2001) Human Cancer Genetics, The Ohio State
University, 420W. 12th Avenue, 606 Medical Research Facility,
Columbus, OH 43210, USA

FEATURES
source
1..347253
location/Qualifiers
1..347253
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="8"
30542..78985
/gene="ATP6C"
/map="8q22"
join(30542..30668,50237..50407,51779..51846,58301..58386,
60481..60575,62162..62253,63315..63413,65299..65367,
72376..72468,72556..72649,74135..74232,75742..75868,
78072..78985)
/gene="ATP6C"
/product="ATPase H+ transporting lysosomal protein"
join(50276..50407,51779..51846,58301..58386,60481..60575,
62162..62253,63315..63413,65299..65367,72376..72468,
72556..72649,74135..74232,75742..75868,78072..78167)
/gene="ATP6C"
/note="vacuolar proton pump; 42kD"
/codon_start=1
/product="ATPase H+ transporting lysosomal protein"
/protein_id="AA150383.1"
/db_xref="GI:17979899"
/translation="MTFFWLISAGEKTCQQTWEKLAATSKNNNAVTSKFNIPDK
VGLIDVAVGLSDELAADAFEGVKKVAQYMDVLEDSKQVENLILANGVDLVYTI
TRFQMDVAKYPIKQSLKISEITAKGTQIDNDLSASAYNNKILQVNDERNGAS
LITRSLAEIVKDDPVLDSRYIVTLIVVVKLHNNDIKQYETLAEVVRSSVLSB
DQDSYLCNVTLPKRAVDDPRHKAREKPFYRDQYVNEEKADKEENRSLSTDKKQF
GPLVRWLKVNPSKAFIWHIVKALRVESVLKRGLPVNFQAMLLQPNKTKLQREV
LHELKRLDSSAAIIDAIPMDIPRLNLSQDEYYPYVYKIDCNLLEK"
78168..78734
/gene="ATP6C"
/note="ATP6C"
/note="U881610; WI-9077; RH62076"
78368..78696
/gene="ATP6C"
/note="SHGC-2582; RH8904; RH101516"
82107..82476
/note="sts-N33243; RH40557"
```

misc\_feature 132102..132445  
/note="cscg48283; RH93096"  
gene 150185..239786  
/gene="BALC"  
/map="8q22.3"  
mRNA join(150185..150527,192841..192945,193785..193931,  
222401..222567,237470..239786)  
/gene="BALC"  
/note="unstable protein product predicted"  
mRNA join(150185..150527,192841..192945,210138..210242,  
222401..222567,237470..239786)  
/gene="BALC"  
/product="BALC 1-2-5-6-8"  
/note="unstable protein product"  
mRNA join(150185..150527,192841..192945,222401..222567,  
237470..239786)  
/gene="BALC"  
/product="BALC 1-2-6-8"  
/note="unstable protein product"  
mRNA join(150185..150527,207578..207648,210138..210242,  
222401..222567,237470..239786)  
/gene="BALC"  
/product="BALC 1-4-5-6-8"  
/evidence=experimental  
mRNA join(150185..150527,210138..210242,222401..222567,  
228946..229035,237470..239786)  
/gene="BALC"  
/product="BALC 1-5-6-7-8"  
/evidence=experimental  
mRNA join(150185..150527,210138..210242,222401..222567,  
237470..239786)  
/gene="BALC"  
/product="BALC 1-5-6-8"  
/evidence=experimental  
mRNA join(150185..150527,237470..239786)  
/gene="BALC"  
/product="BALC 1-8"  
/evidence=experimental  
exon 150185..150527  
/gene="BALC"  
/note="first exon"  
/number=1  
5'UTR /evidence=experimental  
150185..150367  
/gene="BALC"  
/evidence=experimental  
CDS join(150368..150527,210138..210242,222401..222567,  
237470..237580)  
/gene="BALC"  
/note="brain and acute leukemia, cytoplasmic; stable  
protein"  
/codon\_start=1  
/evidence=experimental  
/product="BALC 1-5-6-8"  
/protein\_id="AAL50379.1"  
/db\_xref="GI:17979895"  
/translation="MGCGSRADAIEPRYESWTRETSWLTYTDSAPPSAAPS  
GPEAGLSVLEAEKSKIKAPTDVSDEGLFSAKMAPLAVFSHGMLDGLFSGVPR  
STAPGGINPEKTKNCETQCPNPQSLSGPLTQKONGLOTTEAKDARMPAKEVTIN  
VYDSIQOMDSRRITKNCVN"  
join(150368..150527,222401..222567,237470..237580)  
/gene="BALC"  
/note="brain and acute leukemia, cytoplasmic"  
/codon\_start=1  
/evidence=experimental  
/product="BALC 1-6-8"

CDS /protein\_id="AAL50377.1"  
/db\_xref="GI:17979893"  
/translation="MGCGSRADAIEPRYESWTRETSWLTYTDSAPPSAAPS  
GPEAGLSVLEAEKSKIKAPTDVSDEGLFSAKMAPLAVFSHGMLDGLFSGVPR  
NGLOTTEAKDARMPAKEVTINVTDSIQOMDSRRITKNCVN"  
join(150368..150527,237470..237474)  
/gene="BALC"  
/note="brain and acute leukemia, cytoplasmic"  
/codon\_start=1  
/evidence=experimental  
CDS /product="BALC 1-8"  
/protein\_id="AAL50378.1"  
/db\_xref="GI:17979894"  
/translation="MGCGSRADAIEPRYESWTRETSWLTYTDSAPPSAAPS  
GPEAGLSVLEAEKSKIKAPTDVSDEGLFSAKMAPLAVFSHGMLDGLFSGVPR  
STAPGGINPEKTKNCETQCPNPQSLSGPLTQKONGLOTTEVLLPS"  
join(150368..150527,207578..207639)  
/gene="BALC"  
/note="brain and acute leukemia, cytoplasmic; stable  
protein"  
/codon\_start=1  
/evidence=experimental  
CDS /product="BALC 1-5-6-7-8"  
/protein\_id="AAL50381.1"  
/db\_xref="GI:17979897"  
/translation="MGCGSRADAIEPRYESWTRETSWLTYTDSAPPSAAPS  
GPEAGLSVLEAEKSKIKAPTDVSDEGLFSAKMAPLAVFSHGMLDGLFSGVPR  
STAPGGINPEKTKNCETQCPNPQSLSGPLTQKONGLOTTEVLLPS"  
join(150368..150527,192841..192923)  
/gene="BALC"  
/note="brain and acute leukemia, cytoplasmic; unstable  
protein; encoded by 3 different mRNAs 1-2-6-8, 1-2-5-6-8,  
1-2-3-6-8"  
/codon\_start=1  
/evidence=experimental  
CDS /product="BALC 1-2"  
/protein\_id="AAL50382.1"  
/db\_xref="GI:17979898"  
/translation="MGCGSRADAIEPRYESWTRETSWLTYTDSAPPSAAPS  
GPEAGLSVLEAEKSKIKAPTDVSDEGLFSAKMAPLAVFSHGMLDGLFSGVPR  
STAPGGINPEKTKNCETQCPNPQSLSGPLTQKONGLOTTEVLLPS"  
join(150368..150527,192841..192945)  
/gene="BALC"  
/note="SHGC-101958; RH120941"  
192841..192945  
/gene="BALC"  
/number=2  
/evidence=experimental  
misc\_feature 193785..193931  
/gene="BALC"  
exon /evidence=experimental  
misc\_feature 180807..181082  
/gene="BALC"  
/note="SHGC-147735"  
181419..181834  
/gene="BALC"  
/note="SHGC-101958; RH120941"  
192841..192945  
/gene="BALC"  
/number=2  
/evidence=experimental  
exon /evidence=experimental  
193785..193931  
/gene="BALC"

Alignment Scores:  
Pred. No.: 8.08e+03 Length: 347253  
Score: 52.00 Matches: 11  
Percent Similarity: 85.71% Conservative: 1  
Best Local Similarity: 78.57% Mismatches: 1  
Query Match: 57.78% Indels: 1  
DB: 9 Gaps: 0

US-10-020-441-3 (1-16) x AF363578 (1-347253)



```

gene
complement (4282. .5277)
/ gene="upla"
/ note="XAC3609"
complement (4282. .5277)
/ gene="upla"
/ note="identified by sequence similarity; putative; ORF
located using Blastx/Glimmer/Genemark"
/codon_start=1
/ transl_table=1
/ product="umarylacetate hydrolase"
/ protein_id="AAM38452.1"
/ db_xref="GI:21109988"
/ translation="MMKLGSLKEGGRDGLVAVARLDRTAVRATGAPTLQALAEPM
SNVAPRLNALSINDSADGFDLDUALAPLPRAYEPVDSAYLPHYRVRAN
AVEPESPTDPMQTSAGCYPRDAIKVISEYGDIDLEIYVVDVDMGSPED
AAAHIOLVGLVNDVLRNLIPAEIAGKGFVQSKPRSLSEFVTPDELDAMDSKL
HLPLVTHVNGMFGAPENGMDFADLIHAAKTRPLSGTIVSGSTIANODTTLG
ASCFEORVVEVTLRDKSPTFMAFGDVRIEMFDASGISFGAIEQVERQPLP"
5654. .7036
/ gene="rhlB"
/ note="XAC3610"
5654. .7036
/ gene="rhlB"
/ note="identified by sequence similarity; putative; ORF
located using Blastx/Glimmer/Genemark"
/codon_start=1
/ transl_table=1
/ product="ATP-dependent RNA helicase"
/ protein_id="AAM38453.1"
/ db_xref="GI:21109989"
/ translation="MSFESLGAPFLRLALAEQETPTPIQOQAIPLVLGHDLLAG
AOTGKTAARGLPILOHIGTAPOTVNGRRPALILPTREILATOVHDSKXYL
RIPSAVIYGVGMKQDLALRGVDLIACGRGLDIHERSVDSLSGEVLIIDBAR
MLDGFPSIKRILTKLPQDRQTLFESATRENTKQALSPMRVPOIQVPSVTA
BSITRHPVVGARRKRLDLHLAODSEQTLVARTGSGDKALFLEKSGITVAI
HGNSQCGMRALSDPKAGRYTVLVAIDIAARGDIDQLPVIYNDIPMAEDVHRI
GRTGNSGTGEAISLVAODEAKLRQIVMLGRDVEIIVDGEYEPQPTIRGNSAPGR
AEOPGGDAPRRSHARRPHDAPROAHNAHPKKPGCGRSGPRQANTGAGARRDGS
RGCGRPSGTR"
complement (7238. .9082)
/ gene="XAC3611"
complement (7238. .9082)
/ gene="XAC3611"
/ note="identified by sequence similarity; putative; ORF
located using Blastx/Glimmer/Genemark"
/codon_start=1
/ transl_table=1
/ product="peptidase"
/ protein_id="AAM38454.1"
/ db_xref="GI:21109990"
/ translation="MAATAMSPATWSTSTALPTTKTKRRATAMCARSSIRATARY
GAGSRGRMRSMARSSIRATATFIRPIWDRNWSVQGSVANFATIPGD
PLTPGVGASKGAKRLIEQAGSVLKIPTLPISMGAPQLALAGPVAPBEWRGLPI
TYRIGCDKARVHLKVDADGSOITLYNIATLRGSEYDDQWVRGNHRDGVFGALP
LSGTALLAEAKAIGELAKOGORPKRTLVAVASDSEELGSGTSEAOHDEILAPKA
VLVYTTDNGRGLNAGSHALORLVNSVADVDVDPDGVSLTEFORARADILAPKA
AKPAKDIKTAKRAGGDIPLKALGSGSYSPFLQHLGATTDLGTGGGGSGVTHSL
YDSYDFARFTDPOFAVLPILSQVGRVLRANPVLPRGDFADAVAGAOELKQ
QADSDRTAARTQAEILQAGAYAAVDNPNRPQAPBPKAIVQIDFALDQAITRLQDS

```

Alignment Scores:

Pred. No.:	604	Length:	12514
Score:	51.00	Matches:	2
Percent Similarity:	73.33%	Conservative:	9
Best Local Similarity:	60.00%	Mismatches:	4
Query Match:	56.67%	Indels:	0
DB:	1	Gaps:	0

US-10-020-441-3 (1-16) x AE012011 (1-12514)

Oy 2 GlyTYrGlyArGAspAsnAspArGAspProSeArGlyAsn 16

```

Db 2673 GCGCCGCGACGTGATTCGATCGCGAGATCGTGCAGCCGCAAC 2717
RESULT 23
IMFLCHR34_17/c
Sequence split into 18 fragments LOCUS IMFLCHR34 Accession AL499623
Fragment Name Begin End
IMFLCHR34_00 1 110000
IMFLCHR34_01 100001 210000
IMFLCHR34_02 200001 310000
IMFLCHR34_03 300001 410000
IMFLCHR34_04 400001 510000
IMFLCHR34_05 500001 610000
IMFLCHR34_06 600001 710000
IMFLCHR34_07 700001 810000
IMFLCHR34_08 800001 910000
IMFLCHR34_09 900001 1010000
IMFLCHR34_10 1000001 1110000
IMFLCHR34_11 1100001 1210000
IMFLCHR34_12 1200001 1310000
IMFLCHR34_13 1300001 1410000
IMFLCHR34_14 1400001 1510000
IMFLCHR34_15 1500001 1610000
IMFLCHR34_16 1600001 1710000
IMFLCHR34_17 1700001 1720777
Continuation (18 of 18) of IMFLCHR34 from base 1700001 (AL499623 Leishmania major chromo
Alignment Scores:
Pred. No.: 946
Score: 51.00
Percent Similarity: 70.59%
Best Local Similarity: 64.71%
Query Match: 56.67%
DB: 2
Gaps: 1
Length: 20777
Matches: 11
Conservative: 1
Mismatches: 3
Indels: 2
Oy 2 GlyTYrGlyArGAspAsnAspArGAspProSeArGlyAsn 16
Db 3709 GGAGAGGGCGGATGACCGTGTGATTCGCGACCGCTCGGACAT 3659
RESULT 24
IMFLCHR32_00
COMMENT
Sequence split into 28 fragments LOCUS IMFLCHR32 Accession AL499622
Fragment Name Begin End
IMFLCHR32_00 1 110000
IMFLCHR32_01 100001 210000
IMFLCHR32_02 200001 310000
IMFLCHR32_03 300001 410000
IMFLCHR32_04 400001 510000
IMFLCHR32_05 500001 610000
IMFLCHR32_06 600001 710000
IMFLCHR32_07 700001 810000
IMFLCHR32_08 800001 910000
IMFLCHR32_09 900001 1010000
IMFLCHR32_10 1000001 1110000
IMFLCHR32_11 1100001 1210000
IMFLCHR32_12 1200001 1310000
IMFLCHR32_13 1300001 1410000
IMFLCHR32_14 1400001 1510000
IMFLCHR32_15 1500001 1610000
IMFLCHR32_16 1600001 1710000
IMFLCHR32_17 1700001 1810000
IMFLCHR32_18 1800001 1910000
IMFLCHR32_19 1900001 2010000
IMFLCHR32_20 2000001 2110000
IMFLCHR32_21 2100001 2210000
IMFLCHR32_22 2200001 2310000
IMFLCHR32_23 2300001 2410000
IMFLCHR32_24 2400001 2510000
IMFLCHR32_25 2500001 2610000

```



[illegible]

/product="NADH dehydrogenase subunit 9"  
/protein\_id="BAA9286.1"  
/db\_xref="GI:9049276"  
/translation="MDNQFTFKYSWEPLPKKWKYKIKSEHGNSDNTDYPPOLLCF  
LKHTYTRFOVLIDICVDYPSKRREVEYNLSTYNSKIRLQTCADETRISPV  
SLPFSAGRWEMEDMGVSSINHPDLRLITDGFGBHPLRKDFPGYVEVYDDP  
EKRVSSPIEMTOEFRYFDFAFPWEQRNNEG"  
840. .913  
/product="tRNA-Pro"  
1087. .1160  
/product="tRNA-Tyr"  
/complement(4245. .4712)  
/gene="orf155a"  
/complement(4245. .4712)  
/gene="orf155a"  
/codon\_start=1  
/protein\_id="BAA9287.1"  
/db\_xref="GI:9049277"  
/translation="MSRRGIVEEKTAKSDPIYRNRLVNLVNRILKHGKSLAYQILY  
RAVKIIOQETNPFLSVROAIRGVTPIAVKARVGSTHOVPIEIGSTGKALIR  
MLGARPRGRNNAFKLSSELVAAKSGDAIRKKEETHMAANRAFAHR"  
5954. .6298  
/gene="orf114a"  
5954. .6298  
/gene="orf114a"  
/codon\_start=1  
/protein\_id="BAA9288.1"  
/db\_xref="GI:9049278"  
/translation="MLQCTLRIGIHFLTDPGFRLVGLRDPDDLCCVEGISICSDRL  
RKAGSKRTGSIQRQFFSIIIVFSIRLVVSDPGSVSPFDELLAPAHFVSVDR  
GEGIVGKRIYR"  
complement(6180. .6620)  
/gene="orf146"  
complement(6180. .6620)  
/gene="orf146"  
/codon\_start=1  
/protein\_id="BAA9289.1"  
/db\_xref="GI:9049279"  
/translation="MCFHAPLKPRKMEKMLFLGHTAFVTPPKIMVTPPLTSPMNF  
VIYIEIHYLPKONFELAILPSRQRLTSVEPMHISIRIDMRVQLHCTARICVYLIKMI  
NHASLVQSSSRAPPLLPGRONVGLVTFVHGRDLSRDH"  
7075. .7146  
/product="tRNA-Val"  
complement(10981. .11301)  
/gene="orf106a"  
complement(10981. .11301)  
/gene="orf106a"  
/codon\_start=1  
/protein\_id="BAA9290.1"  
/db\_xref="GI:9049280"  
/translation="MLTCYLTGMMEDGRLVFAWRIGSLRLTGKLSIRPRMWIIS  
LRITIVLSAASSLTILLOTFTLASSPFLILTRINGPGYLLALPVALQRLVLRFLY  
EVLV"  
complement(11066. .12013)  
/gene="orf315"  
complement(11066. .12013)  
/gene="orf315"  
/codon\_start=1  
/protein\_id="BAA9291.1"  
/db\_xref="GI:9049281"  
/translation="WCGKFOIKLYELIRPSAARVHKRIICNNKASPKSFTITWGI  
ONRLYTKDMMANMLACDPVCLGCGAESAOHLFPACEYSOQWGLPFTLRGPPDDC  
FYRLSFWLTGEASFVVGOKTKVPTVYLISLCLOKCAVVDLADSLITVENRND  
QIILAQEKDVAIQELTEFQGYKKLKNSESQULEERKKALBOEVYVTKARSOELD  
AAASATSKATEDFCRSESVADLLFDRYDGGMEACVRCVEDRFPFVIVDAEVDAPQM  
DGHILPKDHPRAVCGCKLPDIYIVANVPELPIPPENDK"  
complement(12617. .12997)  
/gene="orf126"  
complement(12617. .12997)  
/gene="orf126"  
/codon\_start=1  
/protein\_id="BAA9292.1"  
/db\_xref="GI:9049282"

/translation="MSPFLFAIGMEYLSRCLLELTLPNPNVHPRCEKLAITHMFPAS  
DILIFSRADRSVILLYOAFQSSLKPEERAEVYSICEGLRANLKDSSVGLCEAKKAS  
MOEILGIMMCLSIIRLSNSVSP"  
complement(12943. .13332)  
/gene="orf129a"  
complement(12943. .13332)  
/gene="orf129a"  
/codon\_start=1  
/protein\_id="BAA9293.1"  
/db\_xref="GI:9049283"  
/translation="MAVVGCEVINEAOAGRIPEKHIGDNILATELLIKGVSOXAVSPR  
SKIKVDIRKAYDSIEMSFLOCMLSLSEFSPQFTITWMECVTTYSISILNKPFTTTP  
SMKGVQVTPCHSPCLLGWSICLAY"  
complement(14842. .15243)  
/gene="orf133a"  
complement(14842. .15243)  
/gene="orf133a"  
/codon\_start=1  
/protein\_id="BAA9294.1"  
/db\_xref="GI:9049284"  
/translation="MYLARNISRYRADWTRNFEYTPPRTSPFVIRISGHONSRIYRW  
WMEPLIHSOGGMEPQPFVVGILLTKPDAIRSEREEDDILFKGIPGDRPSRLDG  
ISGLGLTGFDLPFFLLERYGNLKSQKSSS"  
complement(18390. .19379)  
/gene="orf329"  
complement(18390. .19379)  
/gene="orf329"  
/codon\_start=1  
/protein\_id="BAA9295.1"  
/db\_xref="GI:9049285"  
/translation="MCLAISLEVYLNPMIAPKPPMAIADRLNTPGKDISLPDLEG  
YVNPYDGGFRLILGKMKHEFPVYFGEKQLSTHKRAEQCKVANTLENQPIINBEY  
LNFLTNDYLVRTGFLPFLNHDIKVAIDIRDSKESERFOKTYTLNSIVHL  
EYNIOARVEONVLEMAVAYRGRIYPAIDRGRIRYTGILHFBHSDLSRLLD  
VMSYGANCYRAETITDCGEDTTHLADETREDDPIIKKULYNNEDLNQHRSY  
AAGGFHYKSYECDAAGAMRIVIKGITQYRARNQVFLSSLDNMLPCQ"  
18651. .19109  
/gene="orf152"  
18651. .19109  
/gene="orf152"  
/codon\_start=1  
/protein\_id="BAA9296.1"  
/db\_xref="GI:9049286"  
/translation="MVKSLITTPYGFSSITVGSITHTNIEKKTPREITLVENQDSS  
AIDTTSVNSGKIYPTKPICSHLQYIILITSPDLVILKYVHDTVASIRLLKITLL  
TTOIADVYCNKIXMVOXRHACGNOIIMSQKIELSLNNKRLVF"  
complement(19539. .19856)  
/gene="orf105a"  
complement(19539. .19856)  
/gene="orf105a"  
/codon\_start=1  
/protein\_id="BAA9297.1"  
/db\_xref="GI:9049287"  
/translation="MRKCWCSGSTCLPSTQRSWLYACDSSGMAIGMKGNRPSSEK  
NRASLMPPTAPAPVPSPKTRIEBSVDSFYQIEISVCRKRVCFNMWSSPLFGLCCVC  
MNT"  
complement(26999. .27343)  
/gene="orf114b"  
complement(26999. .27343)  
/gene="orf114b"  
/codon\_start=1  
/protein\_id="BAA9298.1"  
/db\_xref="GI:9049288"  
/translation="MSRTIGAGCFYNPPLNMWKLOPILPLILOEELHFSLYOSLST  
IKNRTKSIQLEPISGSAEKRESEKLCPIILCSRLGLVQLATPQILKASRSLPYGCEFP  
LLSLTKVNERHS"  
33147. .34880  
/gene="cbs577"  
33147. .34880  
/gene="cbs577"  
/note="orf157"  
/codon\_start=1  
/product="cytochrome c biogenesis protein"

```

/protein_id="BAA99299.1"
/db_xref="GI:9049289"
/translation="MSIYELFHSYLPGLFIATFYTKKOPPAFGAFAFWCILSFLG
LPGHLPNNLSNNILITANAPFPFOISGTSMHSGSLILMCRISGPFGLCYGRSQ
SHNYSKGGHSESLSPFLVLANPVKNISLIPREKRSVLEPOLYTLFVRLTYDSE
LCSRNRTFDGPALEFYVPERKMSFALIGARSRSGRKGKTHPLIARUDKERASS
IDEDRIDALGIAFFPEPLSADDPFRANFVTEPLAESNPVQDPIISAHPCLY
AGDVASAMGFLICRSKMMNGIVALHSPMRDVAKEKNTLLCSAGCVSRIITSEFTL
KFKHVGACVYALLLRSKSLMLLRFRFAFSIMLTALVDGGERAKRPFRRGKK
TTTLPLCWTAGANTVVSDDOPELRICRPLTYGILLGSMWMAHHEIGRGGMWR
DPVENASPMRVATARIHSYILPLHSHTLLINVTPLCCVLGTFGFSIRSGILPVHS
FATDDTIGILFRFPFLMTGISMLFSQMKQASVRRYKKEWVAARSTLVHLHAR
AQPRLMLMKN"
join(36904, .37670,38603, .39152)

```

## Alignment Scores:

```

Pred. No.: 7.03e+03 Length: 199000
Score: 51.00 Matches: 10
Percent Similarity: 71.43% Conservative: 0
Best Local Similarity: 71.43% Mismatches: 4
Query Match: 56.67% Indels: 0
DB: 8 Gaps: 0

```

US-10-020-441-3 (1-16) x AP000396S1 (1-199000)

Oy 3 TyrglyArgAspAspAsnAspArgAspProSerArgLyAsn 16

Db 192809 TATCCGATAGATGATGACGAGGATCCTTACAAAAAT 192850

## RESULT 27

CNS01DVQ/c

LOCUS 201587 bp DNA linear PRI 26-APR-2001  
 DEFINITION Human chromosome 14 DNA sequence BAC R-86092 of library RPCI-11  
 from chromosome 14 of Homo sapiens (Human), complete sequence.

ACCESSION AL136000

VERSION AL136000.4 GI:13872726

KEYWORDS HTG.

SOURCE human.

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE 1 (bases 1 to 201587)

AUTHORS Hellwig, R., Petit, J. L., Vico, V., Dasilva, C., Robert, C., Wincker, P.,

Broctier, P., Caticolico, L., Barbe, V., Pelletier, E., Attiguenave, F.,

Levy, M., Eckenberg, R., Bruls, T., deBardina, V., Crouaud, C.,

Gyapay, G., Saurin, W. and Weissenbach, J.

Sequencing of the human chromosome 14

Unpublished

Genoscope. 2 (bases 1 to 201587)

Direct Submission

Submitted (26-APR-2001) Genoscope - Centre National de Sequencage ;

BP 191 91006 Evry cedex - FRANCE (E-mail : sequef@genoscope.cns.fr)

- Web : www.genoscope.cns.fr

On Apr 27, 2001 this sequence version replaced gi:12140284.

----- Genome Center

Center: Genoscope / Centre National de Sequencage

Center code: GS

Web site: http://www.genoscope.cns.fr/

Contact: Sequef@genoscope.cns.fr

-----

The following BAC sequence is oriented from the T7 to the SP6 end.

Upstream BAC (overlapping the T7 end) : R-543C4

Downstream BAC (overlapping the SP6 end) : R-436M15 (AC=AL133368)

----- Summary Statistics

Assembly program: Phrap; version 2.0

Quality coverage: 5.88x in Q20 bases; sum-of-configs

-----

Overall quality chart :

```

Range : bases
0 : 1
1 - 9 : 10

```

```

10 - 19 : 88
20 - 29 : 264
30 - 39 : 842
40 - 49 : 4368
50 - 59 : 8496
60 - 69 : 14330
70 - 79 : 35308
80 - 89 : 70501
90 - 99 : 67379
-----
Percentage of bases with a quality value >= 40 : 99 %.
Location/Qualifiers
1. .201587
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="14"
/clone="R-860P2"
/clone_id="RPCI-11"
27187. .27362
/notes="matching EMBL:R49568
RHD:RH78065
dbSTS:STS54807
Identified using the e-PCR software (G. Schuler)"
114224. .114397
/notes="matching EMBL:G07703
RHD:RH13195
dbSTS:STS6914
Identified using the e-PCR software (G. Schuler)"

```

## FEATURES

source

STS

STS

BASE COUNT

51162 a 46670 c 46421 g 57334 t

ORIGIN

## Alignment Scores:

```

Pred. No.: 7.11e+03 Length: 201587
Score: 51.00 Matches: 8
Percent Similarity: 91.67% Conservative: 3
Best Local Similarity: 66.67% Mismatches: 1
Query Match: 56.67% Indels: 0
DB: 9 Gaps: 0

```

US-10-020-441-3 (1-16) x CNS01DVQ (1-201587)

Oy 2 gTyrglyArgAspAspAsnAspArgAspProSer 13

Db 53832 GCGTCGGGTAGATGACTCAGACCGACCTGCG 53797

## RESULT 28

AP003012

LOCUS

DEFINITION Mesorhizobium loti DNA, complete genome, section 19/21.

ACCESSION AP003012 BA000012

VERSION AP003012.2 GI:14026998

KEYWORDS Mesorhizobium loti (strain:MAF303099) DNA.

SOURCE

ORGANISM

Mesorhizobium loti

Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;

Phyllobacteriaceae; Mesorhizobium.

REFERENCE

AUTHORS

Kaneko, T., Nakamura, Y., Sato, S., Asamizu, E., Kato, T., Sasamoto, S.,

Watanabe, A., Idesawa, K., Ishikawa, A., Kawashima, K., Kimura, T.,

Kishida, Y., Kiyokawa, C., Kohara, M., Matsumoto, M., Matsuno, A.,

Mochizuki, Y., Nakayama, S., Nakazaki, N., Shimpo, S., Sugimoto, M.,

Takeuchi, C., Yamada, M. and Tabata, S.

Complete genome structure of the nitrogen-fixing symbiotic

bacterium Mesorhizobium loti

DNA Res. 7 (6), 331-338 (2000)

MEDLINE

JOURNAL

AUTHORS

TITLE

REFERENCE

JOURNAL

AUTHORS

TITLE

REFERENCE

JOURNAL

URL: <http://www.kazusa.or.jp/rhizobae/>,  
Tel: 81-438-52-3935 (ex. 2338), Fax: 81-438-52-3934)  
On May 11, 2001 this sequence version replaced gi:11994987.  
location/Qualifiers

source  
1. .346547  
/organism="Mesorhizobium loti"  
/strain="MAF30309"  
/db\_xref="taxon:381"  
complement(4. .507)  
/gene="ml17507"  
complement(4. .507)  
/note="unknown protein"  
/codon\_start=1  
/transl\_table=11  
/protein\_id="BAB53953.1"  
/db\_xref="gi:14026999"  
/translation="MPAFGGIPTMFVDVFWRAVAIGIGATVIMDIMAILLHKAFQPRP  
NMGPGVGMVHLDGKVFHDDIGEAEPVAHEVALGMVFHYFGVIYGIILVLAGTGM,  
AAPFLPAFLIGITVGAGWFLAPGMAGAAKSRPNMIRALNLSHTVFLGLM  
GALLIR"  
complement(574. .1056)  
/gene="ml17508"  
complement(574. .1056)  
/gene="ml17508"  
/codon\_start=1  
/transl\_table=11  
/product="peroxidoxin-like protein"  
/protein\_id="BAB53954.1"  
/db\_xref="gi:14027000"  
/translation="MTISVGDKLPENATFKMTADAKAITSAEIIPGKRVVLFQVPGA  
FTPTCSNNHLEGVLENDAILARGVDITAAVAVDVHMGAMARFTGGEEDKILFLAG  
SADPAKAVGLDNDLSAGMGLRSKRFMSIVDDGVTALNITRKGVDSGAHILIGOL  
"  
complement(1144. .1950)  
/gene="ml17509"  
complement(1144. .1950)  
/gene="ml17509"  
/note="unknown protein"  
/codon\_start=1  
/transl\_table=11  
/protein\_id="BAB53955.1"  
/db\_xref="gi:14027001"  
/translation="MQSLKTIILGAIGASAVLPASASSAMVNSSEGGKVLVTSCKP  
DEAGRIQVDLALKPGWKTYRDPGAGVPOINVSASTINADATLSFPGRHDDG  
YGGAGYVPVSLPVFTLSTAKDSAVIDAVFGICETICIPVOTRLTPRGSDDPN  
ADDAALYKASFAALPAPKASDPGAINVLPGDETLVVEASFPGEDEADFFVAGERDIY  
FGAPSRSEKDKLIFTVPTLDRPSATPTDGLHYLTISAGAVGGLLFP"  
2122. .2730  
/gene="ml17511"  
2122. .2730  
/gene="ml17511"  
/note="hypothetical protein"  
/codon\_start=1  
/transl\_table=11  
/protein\_id="BAB53956.1"  
/db\_xref="gi:14027002"  
/translation="MDLIRHKTKTAAGRGFLDDOFLIAMPKKDDRFTRPSVITGAHSD  
EGAGGLINQOQMLFPDLIVQLGIMNEQELRIPDAQRDVFVRNGGVDSRGFVLH  
SGDTRVESLTVSDICTATVDILIRAISSRGPHALMALGYSGMGQULETAEIN  
GMLTGPASPELLPADIERKYDRILSIGIDLHLSLAAGHA"  
complement(2752. .4416)  
/gene="ml17513"  
complement(2752. .4416)  
/gene="ml17513"  
/note="hypothetical protein"  
/codon\_start=1  
/transl\_table=11  
/protein\_id="BAB53957.1"  
/db\_xref="gi:14027003"  
/translation="WASGSDTVWMDVLRDRVYTKPDVSTIQGLAENS LGGAARNWLP  
VLHADDRTFRTLDVLEHRRGRVAQNFRLGADGHWFALRARPVIGSDGEVTRC  
VGTWVDVTEOKSEBERLLHDAVHNTLQTLPNRLELPMNRLFAIISIA RTEKVRPTVFV

IDIDRKQVNDGIGISAGDTILITIAERHLIKPKDLSRFAGDOPALMLLSBODPA  
RIGVADAIQHAINNPTTPAKREIVLTASGLITWPSAQTSADMDVADALBAQAKR  
FGCDRIEPEFRARTVGTDLRQESDLRRAIRERREFLAQPIVREDGSAVAGEAL  
RMDHPRRGMTPADPFIIVASGGLIVOLGAGQAQAEADLAGQKQGDAPLSVNL  
SSROLIRDLVDSVRGSIARANIKPRCFRELRESLVMNDPEOTAHVLTLOLGIDL  
SLDDFGTGSIALYTRFPDDTKIKDSFPDDMTPRAYVILKSMNNAHLLGSLVAE  
GISDERDALELRQMGCEYVGSFWMGAMPDDQVLTLLKEQYPLTQA"  
complement(5644. .6240)  
/gene="ml17514"  
complement(5644. .6240)  
/gene="ml17514"  
/codon\_start=1  
/transl\_table=11  
/product="ribosomal protein-alanine N-acetyltransferase"  
/protein\_id="BAB53958.1"  
/db\_xref="gi:14027004"  
/translation="MPALPFRFRDLPALKNLVITRVNPNYNDREWSTLRGESAPLE  
PMEKRTPDLDLTAMRLRISRYREDYAGTALAFPIFEKSSKLAGITLGNIRHV  
SQSGHVGWYGERFPGGLMTDVKVAFAPDTLRHLEAACIPDNIRKISIRVLEKA  
GPARREGLRBYLRINGIMQDHYLYARIADPDLDAGTGD"  
complement(6256. .7548)  
/gene="ml17516"  
complement(6256. .7548)  
/gene="ml17516"  
/codon\_start=1  
/transl\_table=11  
/product="processing proteinase"  
/protein\_id="BAB53959.1"  
/db\_xref="gi:14027005"  
/translation="MGTEVSRLSNGLTATETLPSIESVALGMGVSGARNREBEHG  
LAHLLEHMAKGTARTAFIASIEIDVGSINAVTSVETTSYARVLSDDVLAIVI  
LAILQSEBEPQELEREQHVILQIEGAHDPDIDVDFRFTETAFRHQTIIGRIIGT  
PMVVSFSTSKOLDPIEROYGAEKRVVVAAGDIKHDFVAREVEKQLGGFSKADSTIP  
QYAOYVGDFREDRLMDAQIVLGEGRVAHVDVFSOVLMSMLGKMSRLQEVK  
EKRGICYSYVAHSPMGSPSDNGIFGNHATGSDIAELVPIVDELQKGNILQSELRL  
AKQYRAGLIMSASPSASRASQIARQILLGRPIAKELMERLSALTYERLTDLSKM  
FSTKPTLVAVPGVTLAPYEAILDLSGTQVTAARLAV"  
complement(7563. .8960)  
/gene="ml17517"  
complement(7563. .8960)  
/gene="ml17517"  
/codon\_start=1  
/transl\_table=11  
/product="threonine synthase"  
/protein\_id="BAB53960.1"  
/db\_xref="gi:14027006"  
/translation="MQYVSTRGEPALGSPDAVLGARDGGLVYPREMPQFSSEIR  
AMKRLATPDLAIRVLSFPLGSELIPAFVEKLVREAYATFHEAYCPVQGTGNTFVLE  
LFHGPTLAFKDVAMQLLARLMDHVARDDHATVGTSGDTGGAALDAFAGRSTDI  
FILLEPHGRVPPVQORQMTTSKAENVHALATEGNDDCOGLKMFNDHAFDRVALSG  
VNSINMARIMAOIIVYFSSALSGLAPRPVPSFTVGNFQDIFAGYAAKKGGLFERL  
VIATNDNDILARTPATGEVTRKGVAFATTSMDIOTGSNFRLLFEASNDPAATVRY  
MDGLKQSGATFIESREINGRSEFDDARATMDVAAITRSTIASNVILDPHTAAMH  
VAAKASGAPMVVVLGTAHPAKTFAAVEAASGVSPALPWLGGILMTBEKTYVPSDL  
KMVEDYVSRARAR"  
8930. .9817  
/gene="ml17518"  
8930. .9817  
/gene="ml17518"  
/note="unknown protein"  
/codon\_start=1  
/transl\_table=11  
/protein\_id="BAB53961.1"  
/db\_xref="gi:14027007"  
/translation="MLPRAYSHTACLSPPVAWLSVHHKVVNFCAGSSRIPAAALIVYT  
SFTKGSCKPMAFRTLSGRFLAGLALTFPMILAAQCSQDNGILNLGKKDPTAPPP  
QDPKILASQNLNAYCPRYTVRDGTAFNTYANEVKAPKUKTGDAAADAQAQAADA  
AAATPPDPSARIIVQASIVADVTRDCSADQGLAMKIVAGKIVPGPFTPTGTTMPTR  
TAVWGHGVLYSOIHQYOVQVTPDSVATQFVFTDSNVVPAPFTADYQVAVGYDETAP  
KATVDDKTKTHRKAATTN"  
complement(9860. .10549)  
/gene="ml17519"  
complement(9860. .10549)

```
/gene="ml17519"
/feature="hypothesized protein"
/codon_start=1
/transl_table=11
/protein_id="BAB53962.1"
/db_xref="GI:14027008"
/translation="MPQDLVIFPCDDGLVDSSEIIARAVEAEIITLAGYEISAEIAE
SVAGLTFEDIMRIEESKIPFOASLIDRAELVDKRSVRAIEGVRAEASVTAQ
RCICNSRSEIEFMLEKRVHLLPPAGRIFFALRIPSKTKTPADVDVLFPAERIGANP
ANTFVSDSVHGVAGARAAGRVIGFTGASHSYFGHADALTGAQETVIRMAELKSV
IALSEWSADA"
/gene="ml17520"
/feature="ml17520"
/10726..11859
/gene="ml17520"
/10726..11859
/codon_start=1
/transl_table=11
/product="adenine DNA methyltransferase"
/protein_id="BAB53963.1"
/db_xref="GI:14027009"
/translation="MSAVRLHLSHPAQSEMLDTILKGDVVALDLRPEKSVDF
ADPRYNQLDDDLRPPQSKVDVNDHNDQENFEADPAFTRAILARVLRKNGTI
WISSTHNIIFVGAQMQLGFWILNDVWRTNEMPNRFRFQNAHETMIASRDOK
GKGYTFYVLAIRKASNDIQRKDWLFPICGSERLKDNMGDLKPTQPEALIRIM
ASTKPCDILVDPFGSGTTGAVAKRIGRHIFREBOXYIDANERIDAVPLDADL
TVLTKRAEPFVAFVSLIDTGLMTGVTLYDAKKRMAYKADGTVAGDSAGSIHKI
GAEYQGLDAGCGFMWYHRSRGGTLPIDELRIARLGMERAGC"
complement(11901..12539)
```

gene

Alignment Scores:

Pred. No.:	1.15e+04	Length:	346547
Score:	51.00	Matches:	9
Percent Similarity:	83.33%	Conservative:	1
Best Local Similarity:	75.00%	Mismatches:	2
Query Match:	56.67%	Indels:	0
DB:	1	Gaps:	0

US-10-020-441-3 (1-16) x AP003012 (1-346547)

Qy 2 GLYTYRGYARGAPAPAPAPAPAPAPAPROSER 13

Db 235804 GGCTACGGCAGCAGCAATCTCAAGACCCACG 235839

RESULT 29

AF342784/c

LOCUS AF342784 489 bp mRNA linear PLN 26-MAR-2001

DEFINITION Brassica napus GF14 Psia mRNA, partial cds.

ACCESSION AF342784

VERSION AF342784.1

KEYWORDS GI:13447111

SOURCE Brassica napus.

ORGANISM Brassica napus.

REFERENCE

AUTHORS Finemann, J. and Schjoerring, J. K.

TITLE Sequences of Brassica napus regulatory growth factor genes encoding 1-3-3 proteins

JOURNAL Unpublished

AUTHORS Finemann, J. and Schjoerring, J. K.

TITLE Submitted (26-JUN-2001) Plant Nutrition Laboratory, The Royal Veterinary and Agricultural University, Thorvaldsensvej 40, Frederiksberg C, DK-1871, Denmark

FEATURES

source

1..489

/organism="Brassica napus"

/db\_xref="taxon:3708"

<1..>489

/note="1-3-3 protein"

/codon\_start=1

BASE COUNT 131 a 94 c 134 g 130 t

ORIGIN

```
/product="GF14 Psia"
/protein_id="AAK26638.1"
/db_xref="GI:13447112"
/translation="AEQAEYEEVEMERAKSVDELSVEERNLISVAKRVICA
RRASWRIITSSIEQESKGNEDHVAIIKDRGKIEALSKICDILVLEHNLPSAS
PAERKVLNMGDGHRYLAEPKAGARXKDAESTVLAIVKSAODIATVALEPHPIRL
GLA"
```

Alignment Scores:

Pred. No.:	48.5	Length:	489
Score:	50.00	Matches:	9
Percent Similarity:	76.92%	Conservative:	1
Best Local Similarity:	69.23%	Mismatches:	3
Query Match:	55.56%	Indels:	0
DB:	8	Gaps:	0

US-10-020-441-3 (1-16) x AF342784 (1-489)

Qy 4 GLYTAGAPAPAPAPAPAPAPAPROSERXJOLYSAN 16

Db 162 GGAAGAGATGATCTTCACGAGCCCTTCTGCTCCAAAT 124

RESULT 30

AC094913/c

LOCUS AC094913 146970 bp DNA linear HTG 10-JUL-2002

DEFINITION Rattus norvegicus clone CH230-6C20, \*\*\* SEQUENCING IN PROGRESS \*\*\*

ACCESSION AC094913

VERSION AC094913.3

KEYWORDS HTG: HTGS PHASE1.

SOURCE Norway rat.

ORGANISM Rattus norvegicus

REFERENCE

AUTHORS Muzny, D. M., Adams, C., Adio-Oduola, B., Ali-osman, F. R., Allen, C., Alsbrooks, S. L., Amaratunga, H. C., Aze, J. R., Ayele, M., Banks, T., Barbarella, J., Benton, J., Bimaga, K., Blankenhuy, K., Bonnin, D., Bouck, J., Bowie, S., Brieva, M., Brown, E., Brown, M., Bryant, N. P., Buhay, C., Burch, P., Burkett, C., Burrell, K. L., Byrd, N. C., Carton, T. F., Carter, M., Cavazos, S. R., Chacko, J., Chavez, D., Chen, G., Chen, R., Chen, Z., Chowdhry, I., Christopoulos, C., Cleveland, C. D., Cox, C., Coyle, M. D., Dathorne, S. R., David, R., Davila, M. L., Davis, C., Davy-Carroll, L., Dederich, D. A., Delaney, K. R., Delgado, O., Denn, A. L., Ding, Y., Dinh, H. H., Douthwaite, K. J., Draper, H., Dugan-Rocha, S., Durbin, R. J., Earnhart, C., Edgar, D., Edwards, C. C., Elhaj, C., Becotto, M., Falls, T., Ferraguto, D., Flagg, N., Ford, J., Foster, P., Franz, P., Gabisi, A., Gao, J., Garcia, A., Garner, T., Garza, N., Gill, R., Gorrell, J. H., Guevara, W., Gunaratne, P., Hale, S., Hamblin, K., Harris, C., Harris, K., Hart, M., Havlak, P., Hawes, A., Hernandez, J., Hernandez, O., Hodgson, A., Hognes, M., Hollway, J., Hollins, B., Homel, F., Howard, S., Huber, J., Hui, Y. S., Hume, J., Jackson, L. E., Jacobson, B., Jia, Y., Johnson, R., Jolivet, S., Joudan, S., Karlsson, E., Kelly, S., Khan, U., King, L., Koryah, J., Kovar, C., Kratochvic, J., Kureshi, A., Landry, N., Leal, B., Lewis, J. C., Lewis, L., Li, J., Li, Z., Licharge, O., Lieu, C., Liu, J., Liu, W., Loulleged, H., Lozado, R. J., Lu, X., Lucier, A., Lucier, R., Luna, R., Ma, J., Maheshwari, M., Mapa, P., Martin, R., Matlindale, A., Martinez, E., Massey, E., Mawhinney, B., McLeod, M. P., Meador, M., Mel, G., Metzger, M., Miner, G., Miner, Z., Mitchell, T., Mohabbat, K., Morgan, M., Morris, S., Moser, M., Neal, D., Newton, J., Newton, N., Nguyen, A., Nguyen, N., Nguyen, N., Nickerson, S., Nwokenko, S., Ogih, M., Okwona, G., Ogunyemi, N., Oviedo, R., Pace, A., Payton, B., Peery, J., Perez, L., Peters, L., Pickens, R., Primus, E., Pu, L. L., Quiles, M., Ren, Y., Rives, M., Rojas, A., Rojurokan, I., Rolfe, M., Ruiz, S., Savery, G., Scheer, S., Scott, G., Shen, H., Shooshari, N., Sisson, I., Sodergren, E., Sonalke, T., Sparks, A., Stanley, H., Stone, H., Sutton, A., Svatek, A., Tabor, P., Tameirisa, A., Tameirisa, K., Tang, H., Tansey, J., Taylor, C., Taylor, T., Telford, B., Thomas, N., Thomas, S.,

Usmani,K., Vasquez,L., Vera,V., Villalon,D., Vinson,R., Wang,Q.,  
Wang,S., Ward-Moore,S., Warren,R., Washington,C., Watlington,S.,  
Williams,G., Williamson,A., Wleczek,R., Wooden,S., Worley,K.,  
Wu,C., Wu,Y., Wu,Y.F., Zhou,T., Zorrilla,S., Nelson,D.,  
Weinstock,G. and Gibbs,R.  
Direct Submission  
Unpublished  
2 (bases 1 to 146970)  
Worley,K.C.  
Submitted (15-SEP-2001) Human Genome Sequencing Center, Department  
of Molecular and Human Genetics, Baylor College of Medicine, One  
Baylor Plaza, Houston, TX 77030, USA  
3 (bases 1 to 146970)  
Worley,K.C.  
Direct Submission  
Submitted (10-JUL-2002) Human Genome Sequencing Center, Department  
of Molecular and Human Genetics, Baylor College of Medicine, One  
Baylor Plaza, Houston, TX 77030, USA  
On Jul 9, 2002 this sequence version replaced gi:17941710.  
----- Genome Center  
Center: Baylor College of Medicine  
Center code: BCM  
Web site: <http://www.hgsc.bcm.tmc.edu/>  
Contact: [hgsc-help@bcm.tmc.edu](mailto:hgsc-help@bcm.tmc.edu)  
----- Project Information  
Center project name: GBRX  
Center clone name: CH230-6C20  
----- Summary Statistics  
Sequencing vector: Plasmid;  
Chemistry: Dye-terminator Big Dye; 100% of reads  
Assembly program: Phrap; version 0.990329  
Consensus quality: 92271 bases at least Q40  
Consensus quality: 97706 bases at least Q10  
Consensus quality: 102314 bases at least Q20  
-----  
\* NOTE: Estimated insert size may differ from sequence length  
\* (see [http://www.hgsc.bcm.tmc.edu/docs/Genbank\\_draft\\_data.html](http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html)).  
\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 59 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved.  
1 1179: contig of 1179 bp in length  
\* 1180 1279: gap of unknown length  
\* 1280 2496: contig of 1217 bp in length  
\* 2497 2596: gap of unknown length  
\* 2597 4170: contig of 1574 bp in length  
\* 4171 4270: gap of unknown length  
\* 4271 5575: contig of 1305 bp in length  
\* 5576 5675: gap of unknown length  
\* 5676 7019: contig of 1344 bp in length  
\* 7020 7120 7119: gap of unknown length  
\* 7120 8691: contig of 1572 bp in length  
\* 8692 8791: gap of unknown length  
\* 8792 10183: contig of 1382 bp in length  
\* 10184 10283: gap of unknown length  
\* 10284 11471: contig of 1188 bp in length  
\* 11472 11571: gap of unknown length  
\* 11572 12723: contig of 1152 bp in length  
\* 12724 12823: gap of unknown length  
\* 12824 14403: contig of 1580 bp in length  
\* 14404 14503: gap of unknown length  
\* 14504 15527: contig of 1024 bp in length  
\* 15528 15627: gap of unknown length  
\* 15628 17611: contig of 1984 bp in length  
\* 17612 17711: gap of unknown length  
\* 17712 19429: contig of 1718 bp in length  
\* 19430 19529: gap of unknown length  
\* 19530 21227: contig of 1698 bp in length

21228 21327: gap of unknown length  
\* 21328 22376: contig of 1049 bp in length  
\* 22377 22476: gap of unknown length  
\* 22477 24009: contig of 1533 bp in length  
\* 24010 24109: gap of unknown length  
\* 24110 26423: contig of 2314 bp in length  
\* 26424 26523: gap of unknown length  
\* 26524 27957: contig of 1434 bp in length  
\* 27958 28057: gap of unknown length  
\* 28058 29638: contig of 1581 bp in length  
\* 29639 29738: gap of unknown length  
\* 29739 31846: gap of unknown length  
\* 31847 31946: gap of unknown length  
\* 31947 34142: contig of 2196 bp in length  
\* 34143 34242: gap of unknown length  
\* 34243 36376: contig of 2134 bp in length  
\* 36377 36476: gap of unknown length  
\* 36477 38561: contig of 2085 bp in length  
\* 38562 38661: gap of unknown length  
\* 38662 39903: contig of 1242 bp in length  
\* 39904 40003: gap of unknown length  
\* 40004 41958: contig of 1955 bp in length  
\* 41959 42058: gap of unknown length  
\* 42059 43625: contig of 1567 bp in length  
\* 43626 43725: gap of unknown length  
\* 43726 45011: contig of 1286 bp in length  
\* 45012 45111: gap of unknown length  
\* 45112 47277: contig of 2166 bp in length  
\* 47278 47377: gap of unknown length  
\* 47378 49421: contig of 2044 bp in length  
\* 49422 49521: gap of unknown length  
\* 49522 51042: contig of 1521 bp in length  
\* 51043 51142: gap of unknown length  
\* 51143 53482: contig of 2340 bp in length  
\* 53483 53582: gap of unknown length  
\* 53583 55075: contig of 1493 bp in length  
\* 55076 55175: gap of unknown length  
\* 55176 57330: contig of 2155 bp in length  
\* 57331 57430: gap of unknown length  
\* 57431 60307: contig of 2877 bp in length  
\* 60308 60407: gap of unknown length  
\* 60408 62705: contig of 2298 bp in length  
\* 62706 62805: gap of unknown length  
\* 62806 64827: contig of 2022 bp in length  
\* 64828 64927: gap of unknown length  
\* 64928 67634: contig of 2707 bp in length  
\* 67635 67734: gap of unknown length  
\* 67735 69903: contig of 2169 bp in length  
\* 69904 70003: gap of unknown length  
\* 70004 72964: contig of 2961 bp in length  
\* 72965 73064: gap of unknown length  
\* 73065 75860: contig of 2796 bp in length  
\* 75861 75960: gap of unknown length  
\* 75961 77903: contig of 1943 bp in length  
\* 77904 78003: gap of unknown length  
\* 78004 80744: contig of 2740 bp in length  
\* 80744 80844 83862: contig of 3019 bp in length  
\* 83863 83962: gap of unknown length  
\* 83963 86802: contig of 2840 bp in length  
\* 86803 86902: gap of unknown length  
\* 86903 90056: contig of 3154 bp in length  
\* 90057 90156: gap of unknown length  
\* 90157 92841: contig of 2685 bp in length  
\* 92842 92941: gap of unknown length  
\* 92942 94466: contig of 1525 bp in length  
\* 94467 94566: gap of unknown length  
\* 94567 97310: contig of 2744 bp in length  
\* 97311 97410: gap of unknown length  
\* 97411 99793: contig of 2383 bp in length  
\* 99794 99893: gap of unknown length  
\* 99894 103484: contig of 3591 bp in length  
\* 103485 103584: gap of unknown length

\* 103585 106611: contig of 3027 bp in length  
\* 106612 106711: gap of unknown length  
\* 106712 110006: contig of 3395 bp in length  
\* 110007 110106: gap of unknown length  
\* 110107 113951: contig of 3845 bp in length  
\* 113952 114051: gap of unknown length  
\* 114052 118104: contig of 4053 bp in length

Alignment Scores:  
Pred. No.: 7.66e+03 Length: 146970  
Score: 50.00 Matches: 8  
Percent Similarity: 84.62% Conservative: 3  
Best Local Similarity: 61.54% Mismatches: 2  
Query Match: 55.56% Indels: 0  
DB: Gaps: 0

US-10-020-441-3 (1-16) x AC094913 (1-146970)

Qy 1 ValgIYrGlyArgAspAspAsnAspArgAspProSer 13  
Db 34839 GTTCATATACCAAAATGACATGACAAAGTCTTCA 34801

RESULT 31  
AC129726  
LOCUS Rattus norvegicus clone CH230-383J24, \*\*\* SEQUENCING IN PROGRESS  
DEFINITION \*\*\* 45 unordered pieces.  
AC129726 174255 bp DNA linear HTG 02-AUG-2002  
VERSION AC129726.1 GI:22091219  
KEYWORDS HTG; HTGS PHASE1.  
SOURCE Norway rat.  
ORGANISM Rattus norvegicus  
Eukaryote; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;  
Rattus.  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;  
Rattus.  
1 (bases 1 to 174255)  
Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C.,  
Aleshrooke,S.L., Amarantunge,H.C., Are,J.R., Ayale,M., Banks,T.,  
Barbata,J., Benton,J., Bimaga,K., Blankenburg,K., Bonnin,D.,  
Bouck,J., Bowie,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P.,  
Buhay,C., Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C.,  
Carron,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D.,  
Chen,G., Chen,R., Chen,Z., Chowdhury,I., Christopoulos,C.,  
Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R.,  
Devila,M.L., Davis,C., Davy-Carroll,L., Dederich,D.A.,  
DeLaney,K.R., Delgado,O., Dem,A.L., Ding,Y., Dim,H.H.,  
Douthwaite,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.J.,  
Eamhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Escotto,M.,  
Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Frantz,P.,  
Gorrel,J., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R.,  
Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S., Hamilton,K.,  
Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A., Hernandez,J.,  
Hennahad,O., Hodgson,A., Hogue,M., Holloway,C., Hollins,B.,  
Homesi,F., Howard,S., Huber,J., Hulik,S., Hume,J., Jackson,L.E.,  
Jacobson,B., Jia,Y., Johnson,R., Jolivet,S., Joudah,S.,  
Karleson,B., Kelly,S., Khan,U., King,L., Korvah,J., Kovar,C.,  
Kratovic,J., Kureshi,A., Landy,N., Leal,B., Lewis,L.C., Lewis,L.,  
Li,J., Li,Z., Lichteberg,O., Lieu,C., Liu,J., Liu,W., Louisged,H.,  
Lozad,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R., Ma,J.,  
Maheshwari,M., Mapa,P., Martin,R., Martindale,A., Martinez,E.,  
Massey,E., Mawhney,E., McLeod,M.P., Meador,M., Mei,G., Metzger,M.,  
Miner,G., Miner,Z., Mitchell,T., Mohabbat,K., Morgan,M., Morris,S.,  
Moser,M., Neal,D., Newton,J., Newton,N., Nguyen,A., Nguyen,N.,  
Nguyen,N., Nickerson,E., Nwokwenkwo,S., Ogun,W., Okunodu,G.,  
Otagunye,N., Oviado,R., Pace,A., Payton,B., Peery,J., Perez,L.,  
Peterson,L., Pickens,R., Primus,E., Pu,L.L., Quiles,M., Ren,Y.,  
Rives,M., Rojas,A., Rojibokan,I., Rolfe,M., Ruiz,S., Savary,G.,  
Schere,S., Scott,G., Shen,H., Shoohtari,N., Sisson,I.,  
Sodergren,E., Sonalke,T., Sparks,A., Stanley,H., Stone,H.,  
Sutton,A., Svatek,A., Tabor,P., Tamerias,A., Tamerisa,K., Tang,H.,  
Tansey,J., Taylor,C., Taylor,T., Telford,B., Thomas,N., Thomas,S.,  
Umanan,K., Vasquez,L., Vera,V., Villalón,D., Vinson,R., Wang,Q.,  
Wang,S., Ward-Moore,S., Warren,R., Washington,C., Watlington,S.,

TITLE  
JOURNAL  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
COMMENT  
Center: Baylor College of Medicine  
Center code: BCM  
Web site: http://www.hgsc.bcm.tmc.edu/  
Contact: hgsc-help@bcm.tmc.edu  
Project Information  
Center project name: KAHK  
Center clone name: CH230-383J24  
Sequencing vector: Plasmid  
Chemistry: Dye-terminator Big Dye 100% of reads  
Assembly program: Phrap, version 0.990329  
Consensus quality: 131956 bases at least Q40  
Consensus quality: 138984 bases at least Q30  
Consensus quality: 143260 bases at least Q20  
NOTE: Estimated insert size may differ from sequence length  
(see http://www.hgsc.bcm.tmc.edu/docs/Gendbankdraft\_data.html).  
NOTE: This is a 'working draft' sequence. It currently  
consists of 45 contigs. The true order of the pieces  
is not known and their order in this sequence record is  
arbitrary. Gaps between the contigs are represented as  
runs of N, but the exact sizes of the gaps are unknown.  
This record will be updated with the finished sequence  
as soon as it is available and the accession number will  
be preserved.  
1 1630: contig of 1630 bp in length  
1730: gap of unknown length  
1731 2735: contig of 1005 bp in length  
2736 2835: gap of unknown length  
2836 3845: contig of 1010 bp in length  
3846 3945: gap of unknown length  
3946 5405: contig of 1460 bp in length  
5406 5505: gap of unknown length  
5506 7321: contig of 1816 bp in length  
7322 7421: gap of unknown length  
7422 8659: contig of 1238 bp in length  
8660 8759: gap of unknown length  
8760 10011: contig of 1252 bp in length  
10012 10111: gap of unknown length  
10112 11587: contig of 1476 bp in length  
11588 11687: gap of unknown length  
11688 12861: contig of 1174 bp in length  
12862 12961: gap of unknown length  
12962 14037: contig of 1071 bp in length  
14038 14132: gap of unknown length  
14133 15263: contig of 1131 bp in length  
15264 15363: gap of unknown length  
15364 17909: contig of 2546 bp in length  
17910 18009: gap of unknown length  
18010 20129: contig of 2120 bp in length  
20130 20229: gap of unknown length  
20230 22615: contig of 2386 bp in length  
22616 22715: gap of unknown length  
22716 25021: contig of 2306 bp in length  
25022 25121: gap of unknown length  
25122 27038: contig of 1917 bp in length  
27039 27138: gap of unknown length  
27139 29595: contig of 2457 bp in length  
29596 30802: gap of unknown length  
30803 30902: contig of 1107 bp in length  
30903 30902: gap of unknown length

[illegible]

```

Db 141768 ATVAGAGAAGGAAGATGATGATGAGACACGAGGACTGCTGAGAGGAAC 141815

RESULT 32
AC113361 91276 bp DNA linear HTG 13-MAR-2002
LOCUS AC113361
DEFINITION Homo sapiens chromosome 5 clone CTD-3224K15, WORKING DRAFT
SEQUENCE, 3 ordered pieces.
AC113361
AC113361.2 GI:19387600
AC113361.1 HTGS_PHASE2; HTGS_DRAFT; HTGS_ACTIVEFIN.
VERSION AC113361.2
KEYWORDS Homo sapiens.
SOURCE Homo sapiens.
ORGANISM Homo sapiens.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE
AUTHORS DOE Joint Genome Institute.
TITLE Sequencing of Human Chromosome 5
JOURNAL Unpublished
PAGES 2 (bases 1 to 91276)
DOE Joint Genome Institute.
DIRECT SUBMISSION Submitted (01-MAR-2002) Production Sequencing Facility, DOE Joint
Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
3 (bases 1 to 91276)
DOE Joint Genome Institute.
DIRECT SUBMISSION Submitted (13-MAR-2002) Production Sequencing Facility, DOE Joint
Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
On Mar 13, 2002 this sequence version replaced gi:19033475.
-----Genome Center
Center: Joint Genome Institute
Center Code: JGI
Web site: http://www.jgi.doe.gov

-----
Project Information
Center Project Name: 942300
Center clone name: CITB-El_3224K15

-----
Summary Statistics
Consensus quality: 89384 bases at least Q40
Consensus quality: 90516 bases at least Q30
Consensus quality: 90668 bases at least Q20
Estimated insert size: 97000; agarose-fp estimation
Estimated insert size: 91076; sum-of-contigs estimation
Quality coverage: 10.84 in Q20 bases; agarose-fp estimation
Quality coverage: 11.54 in Q20 bases; sum-of-contigs estimation.
* NOTE: This is a 'working draft' sequence. It currently
* consists of 3 contigs. Gaps between the contigs
* are represented as runs of N. The order of the pieces
* is believed to be correct as given, however the sizes
* of the gaps between them are based on estimates that have
* provided by the submitter.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.
* 1 35158: contig of 35158 bp in length
* 35159 35258: gap of unknown length
* 35259 55749: contig of 20491 bp in length
* 55750 55849: gap of unknown length
* 55850 91276: contig of 35427 bp in length.

Location/Qualifiers
1. 91276
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="5"
/clone="CTD-3224K15"
/clone_lib="Caltech human BAC library D"
BASE COUNT 19215 a 24668 c 25288 g 21905 t 200 others
ORIGIN

Alignment Scores:
Pred. No.: 7.16e+03 Length: 91276
Score: 49.00 Matches: 8

```









13637 14826: contig of 1190 bp in length  
14827 14926: gap of unknown length  
14927 16368: contig of 1442 bp in length  
16369 16468: gap of unknown length  
16469 18003: contig of 1535 bp in length  
18004 18103: gap of unknown length  
18104 19279: contig of 1176 bp in length  
19280 19379: gap of unknown length  
19380 20559: contig of 1180 bp in length  
20560 20659: gap of unknown length  
20660 22270: contig of 1611 bp in length  
22271 22370: gap of unknown length  
22371 24113: contig of 1743 bp in length  
24114 24213: gap of unknown length  
24214 25370: contig of 1157 bp in length  
25371 25470: gap of unknown length  
25471 27443: contig of 1973 bp in length  
27444 27543: gap of unknown length  
27544 28651: contig of 1108 bp in length  
28652 28751: gap of unknown length  
28752 29835: contig of 1084 bp in length  
29836 29935: gap of unknown length  
29936 31789: contig of 1854 bp in length  
31790 31889: gap of unknown length  
31890 33078: contig of 1189 bp in length  
33079 33178: gap of unknown length  
33179 34969: contig of 1791 bp in length  
34970 35069: gap of unknown length  
35070 36921: contig of 1852 bp in length  
36922 37021: gap of unknown length  
37022 38441: contig of 1420 bp in length  
38442 38541: gap of unknown length  
38542 40096: contig of 1555 bp in length  
40097 40196: gap of unknown length  
40197 42426: contig of 2230 bp in length  
42427 42526: gap of unknown length  
42527 43801: contig of 1275 bp in length  
43802 43901: gap of unknown length  
43902 45747: contig of 1846 bp in length  
45748 45847: gap of unknown length  
45848 46978: contig of 1131 bp in length  
46979 47078: gap of unknown length  
47079 48253: contig of 1175 bp in length  
48254 48353: gap of unknown length  
48354 49372: contig of 1019 bp in length  
49373 49472: gap of unknown length  
49473 50916: contig of 1444 bp in length  
50917 51016: gap of unknown length  
51017 53199: contig of 2183 bp in length  
53200 53299: gap of unknown length  
53300 54977: contig of 1678 bp in length  
54978 55077: gap of unknown length  
55078 56925: contig of 1848 bp in length  
56926 57025: gap of unknown length  
57026 58598: contig of 1573 bp in length  
58599 58698: gap of unknown length  
58699 60564: contig of 1866 bp in length  
60565 60664: gap of unknown length  
60665 62988: contig of 2324 bp in length  
62989 63088: gap of unknown length  
63089 64818: contig of 1730 bp in length  
64819 64918: gap of unknown length  
64919 66295: contig of 1377 bp in length  
66296 66395: gap of unknown length  
66396 67874: contig of 1479 bp in length  
67875 67974: gap of unknown length  
67975 70300: contig of 2326 bp in length  
70301 70400: gap of unknown length  
70401 73320: contig of 2820 bp in length  
73321 73320: gap of unknown length  
73321 75489: contig of 2169 bp in length  
75490 75589: gap of unknown length  
75590 77717: contig of 2128 bp in length

77718 77817: gap of unknown length  
77818 80449: contig of 2632 bp in length  
80450 80549: gap of unknown length  
80550 82457: contig of 1908 bp in length  
82458 82557: gap of unknown length  
82558 84901: contig of 2344 bp in length  
84902 85001: gap of unknown length  
85002 86747: contig of 1746 bp in length  
86748 86847: gap of unknown length  
86848 89413: contig of 2566 bp in length  
89414 89513: gap of unknown length  
89514 92860: contig of 3347 bp in length  
92861 92960: gap of unknown length  
92961 94865: contig of 1905 bp in length

Alignment Scores:  
Pred. No.: 1.19e+04 Length: 161188  
Score: 49.00 Matches: 9  
Percent Similarity: 75.00% Conservative: 3  
Best Local Similarity: 56.25% Mismatches: 4  
Query Match: 54.44% Indels: 0  
DB: 2 Gaps: 0

US-10-020-441-3 (1-16) x AC095827 (1-161188)  
Qy 1 ValGIYrGIYArgAspAsnAParGAPProSerArgLYsAsn 16  
Db 4896 GTGGGGTGGGGAGAGAGAGAGGAGACAGATCCCTTAAGGGGTAAc 4849

RESULT 36  
AC108127/c  
LOCUS  
DEFINITION Homo sapiens chromosome 5 clone RP11-68E23, WORKING DRAFT SEQUENCE,  
24 unordered pieces.  
ACCESSION AC108127.1 GI:18369973  
VERSION AC108127.1  
KEYWORDS HTG: HTGS PHASE1; HTGS\_DRAFT; HTGS\_ACTIVEFLN.  
SOURCE Homo sapiens.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
REFERENCE 1 (bases 1 to 195653)  
AUTHORS DOE Joint Genome Institute.  
TITLE Sequencing of Human Chromosome 5  
JOURNAL Unpublished  
REFERENCE 2 (bases 1 to 195653)  
AUTHORS DOE Joint Genome Institute.  
TITLE Direct Submission  
JOURNAL Submitted (25-JAN-2002) Production Sequencing Facility, DOE Joint  
Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA  
COMMENT -----Genome Center  
Center: Joint Genome Institute  
Center Code: JGI  
Web site: http://www.jgi.doe.gov  
-----  
Project Information  
Center Project Name: 432596  
Center clone name: RPC1-11\_68E23  
-----  
Summary Statistics  
Consensus quality: 171359 bases at least Q40  
Consensus quality: 184001 bases at least Q30  
Consensus quality: 186322 bases at least Q20  
Estimated insert size: 192000; agarose-fp estimation  
Estimated insert size: 193353; sum-of-contigs estimation  
Quality coverage: 5.25 in Q20 bases; agarose-fp estimation  
Quality coverage: 5.21 in Q20 bases; sum-of-contigs estimation.  
\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 24 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence



TITLE  
JOURNAL

REFERENCE  
AUTHORS

Mccarthy,M., McEwan,P., McKernan,K., Mphahlela,R., Meldrum,J.,  
Mencus,L., Mihova,T., Mienna,V., Murphy,T., Naylor,J., Nguyen,C.,  
Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neill,D.,  
Oliver,J., Peterson,K., Phunkhang,P., Pierre,N., Pollara,V.,  
Raymond,C., Retta,R., Rieback,B., Riley,R., Rise,C., Rogov,P.,  
Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,S., Schuppack,R.,  
Seaman,S., Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,  
Strassus,N., Subramanian,A., Talamas,J., Tsafaye,S., Theodore,J.,  
Topham,K., Travers,M., Travis,N., Trigilio,J., Vassiliou,H.,  
Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,  
Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.

Submitted (23-NOV-2001) Whitehead Institute/MIT Center for Genome  
Research, 320 Charles Street, Cambridge, MA 02141, USA  
3 (bases 1 to 214792)

Birten,B., Nusbaum,C., Lander,E., Ali,A., Allen,N., Anderson,S.,  
Barna,N., Bastien,V., Bloom,T., Boguslavsky,L., Boukhalter,B.,  
Camarata,J., Chang,J., Chazaro,K., Chopel,Y., Collymore,A.,  
Cook,A., Cooke,P., DeArrelano,B., Dewar,K., Diaz,J.S., Dodge,S.,  
Faro,S., Ferreira,P., Fitzgerald,M., Gage,D., Galagan,J.,  
Gardyna,S., Goid,S., Graham,L., Grand-Pierre,N., Hagos,B.,  
Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C., Kanat,A.,  
Karatas,A., Kellis,C., Landert,T., Levine,R., Linblad-Toh,K.,  
Liu,G., Maclean,C., Macdonald,P., Major,J., Matthews,C.,  
Tsafaye,S., Theodore,J., Topham,K., Travers,M., Vassiliev,H.,  
Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Young,G., Zainoun,J.,  
Zembek,L., Zimmer,A. and Zody,M.

Direct Submission  
Submitted (31-AUG-2002) Whitehead Institute/MIT Center for Genome  
Research, 320 Charles Street, Cambridge, MA 02141, USA  
On Aug 21, 2002 this sequence version replaced gi:1706117.  
All repeats were identified using RepeatMasker:  
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>

COMMENT

TITLE  
JOURNAL

Genome Center

Center: Whitehead Institute/ MIT Center for Genome Research

Center code: WIBR

Web site: <http://www-seq.wi.mit.edu>

Contact: [sequence\\_submissions@genome.wi.mit.edu](mailto:sequence_submissions@genome.wi.mit.edu)

Project Information

Center project name: L18028

Center clone name: 9 K 20

----- Summary Statistics -----

Sequencing vector: plasmid; n/a; 100% of reads

Chemistry: Dye-terminator Big Dye; 100% of reads

Assembly program: Phrap; version 0.960731

Consensus quality: 21009 bases at least Q40

Consensus quality: 212426 bases at least Q30

Consensus quality: 212988 bases at least Q20

Insert size: 22000; agarose-fp

Insert size: 21392; sum-of-contigs

Quality coverage: 7.5 in Q20 bases; agarose-fp

Quality coverage: 7.8 in Q20 bases; sum-of-contigs

----- NOTE: This is a 'working draft' sequence. It currently consists of 16 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown. This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved.

\* \* \* \* \*

I 16242: contig of 16242 bp in length

\* 16243 16342: gap of 100 bp

\* 16343 16975: contig of 633 bp in length

\* 16976 17075: gap of 100 bp

\* 17076 17875: contig of 800 bp in length

\* 17876 17975: gap of 100 bp  
\* 17976 19289: contig of 1314 bp in length  
\* 19290 19389: gap of 100 bp  
\* 19390 49754: contig of 30365 bp in length  
\* 49755 49854: gap of 100 bp  
\* 49855 54499: contig of 4645 bp in length  
\* 54500 54599: gap of 100 bp  
\* 54600 59792: contig of 5193 bp in length  
\* 59793 59892: gap of 100 bp  
\* 59893 67895: contig of 8003 bp in length  
\* 67896 67995: gap of 100 bp  
\* 67996 77109: contig of 9114 bp in length  
\* 77110 77209: gap of 100 bp  
\* 77210 87565: contig of 10356 bp in length  
\* 87566 87665: gap of 100 bp  
\* 87666 99932: contig of 12267 bp in length  
\* 99933 100032: gap of 100 bp  
\* 100033 11688: contig of 11656 bp in length  
\* 11689 11788: gap of 100 bp  
\* 11789 140973: contig of 29185 bp in length  
\* 140974 141073: gap of 100 bp  
\* 141074 180265: contig of 39192 bp in length  
\* 180266 180365: gap of 100 bp  
\* 180366 211947: contig of 31582 bp in length  
\* 211948 212047: gap of 100 bp  
\* 212048 214792: contig of 2745 bp in length.

FEATURES  
source  
1. 214792  
/organism="Mus musculus"  
/db\_xref="taxon:10090"  
/clone="RP23-9K20"  
/clone\_1fb="RPC1-23 Female Mouse BAC"  
1. 16242  
/note="assembly\_fragment  
clone\_end:SP6  
vector\_side:left"  
16343. 16975  
/note="assembly\_fragment"  
17076. 17875  
/note="assembly\_fragment"  
17976. 19289  
/note="assembly\_fragment"  
19390. 49754  
/note="assembly\_fragment"  
49855. 54499  
/note="assembly\_fragment"  
54600. 59792  
/note="assembly\_fragment"  
59893. 67895  
/note="assembly\_fragment"  
67996. 77109  
/note="assembly\_fragment"  
77210. 87565  
/note="assembly\_fragment"  
87666. 99932  
/note="assembly\_fragment"  
100033. 11688  
/note="assembly\_fragment"  
11789. 140973  
/note="assembly\_fragment"  
141074. 180265  
/note="assembly\_fragment"  
180366. 211947  
/note="assembly\_fragment"  
212048. 214792  
/note="assembly\_fragment  
clone\_end:T7  
vector\_side:right"

BASE COUNT 64594 a 41827 c 41629 g 65238 t 1504 others  
ORIGIN  
Alignment Scores:  
Pred. No.: 1.53e+04 Length: 214792

Score: 49.00 Matches: 9  
Percent Similarity: 71.43% Conservative: 1  
Best Local Similarity: 64.29% Mismatches: 4  
Query Match: 54.44% Indels: 0  
DB: 2 Gaps: 0

US-10-020-441-3 (1-16) x AC102091 (1-214792)

Oy 3 TyrGIYArgAspAspAsnAPtArgAspProSerArgLYsAsn 16

Db 119336 TATGAAAGGGCTGAAAGAGAGATCTCTTAGGAAATAAT 119377

RESULT 39

SC9718/c 35811 bp DNA linear PLN 11-AUG-1997

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

Notes:

FEATURES

source

CDS

1. 35811  
/organism="Saccharomyces cerevisiae"  
/strain="AB972"  
/db\_xref="taxon:4932"  
/chromosome="XIII"  
/map="13R"  
/clone="cosmid 9718"  
complement<1. 1537  
/note="YM9718.01c, unknown, len: >512, CAI: 0.14, similar  
to SW:YKML YEAST P32330 hypothetical Tvp-Asp repeats  
containing protein in YPK1-SRP21 intergenic region  
(YK121W or YK1525) (49.1% identity in 426 aa overlap)"  
/codon\_start=1  
/product="unknown"  
/protein\_id="CAA89738.1"  
/db\_xref="GI:817860"  
/db\_xref="SWISS-PROT:O03177"  
/translation="MTNMAADKTSISVHSEFSNSTGSEERSRHQSPGSSS  
GESKSKLNDEVSADIEPLKFRNTKTNNDKLSNNGNATDSFRLKHLDORNT  
LNSNLRVNEFPENSIDTEQFENLRPEPKYIKMLKRKNLKKQFRLFLAQLDLVAEES  
TVTSTKSSEPTSKAIVSTKFSRDKGMATGSKGKIRIKVIGSSPVERAELEDSAS  
NKEARAKSMRIKQOVSSLNPKKEQFLDSATEKYEKEKLLNLVAPVHFHPLRLYKE

CDS  
/complement(2087..2449)  
/note="YM9718.02c, questionable orf, len: 120, CAI: 0.07"  
/codon\_start=1  
/product="unknown"  
/protein\_id="CA89739.1"  
/db\_xref="gi:817861"  
/translation="MVRGCALLATLRLMLHFEYSIRILPSTELRSKTDLLSA  
EADRLPADYAKAGAKRSRLPEFAKQRTIALPERKDLFLPLFLHSGRELIS  
SADPPGFCAGASAMGSLT"  
complement(2605..4638)  
/gene="YPK2"  
complement(2605..4638)  
/note="YM9718.03c, YPK2 gene (YKR2), len: 677, CAI: 0.13,  
SW:YPK2 YEAST P18961, serine/threonine-protein kinase  
YPK2/YKR2"  
/codon\_start=1  
/product="ypk2p"  
/protein\_id="CA89740.1"  
/db\_xref="gi:817862"  
/translation="MHSWRISKFLGRKEDGSSSEDENEKMGNGLFPHHGKHH  
GSPKNNHEHHIRKINTNETLPSLSNDKASPKNSGIGNDKSKERKASQ  
SSTEQGSSSESGATVYKSGDFTLPFTKSDSTLIQKLSGILTSSNDSEVA  
AIMKQLPRYKQVQDSAGEGLIDRAFAKFLPSLILFGSTNSSGLFTTEPNSIT  
TISPMGTMEQPVNFKSTFDVTRKLARKLIDVAPLPSLILPSKMOQELGEDEVLT  
KEIKIINTNODIHLSDFHLPLNKLIDSAQIRLYNHMHISLERGYKGLNTVYKPS  
KNKSLSIDDPDLKLVIGSGRKYQVQKQKOTKLYALKARAYVSKCVTTIAE  
RTYLARDCEPTVPLKRSFQSPBEKLYLVLATINGSELFYHIOHGRSLASRYTIAE  
LPLCADLSLHKLDVLYRDLKPENILLDYQGHALCDPGLCKLMKNDKDTDFCGPEY  
LAPBILLGQGYKTVDWMTLGLILYEMWTKGLPYVDENVPMYKKILQOPLLPDGPD  
PAKDLILGLSRDPSRLGVNGDEIRNHFPGDISWKLLKGYIPPRKIVYSEI  
DTANPDQFTKEKPIDSVDEYLSAISIQKPGWMTYIGDQLGDSPGSGSIS"  
complement(5082..6791)  
/gene="PGM2"  
complement(5082..6791)  
/note="YM9718.04c, PGM2 gene, (GAL5) SW:PGM2 YEAST P37012  
phosphoglucosyltransferase 2, PS00710 Phosphoglucosyltransferase and  
phosphomannomutase phosphoserine signature"  
/codon\_start=1  
/product="pgm2p"  
/protein\_id="CA89741.1"  
/db\_xref="gi:817863"  
/translation="MSRQETVPTKPYEDOKRGTSGLRKTKVPEDEPNYENFIQSI  
MEATPEGSKATLVVGGDGRYYNDVILKINATIGANGIKLVIGQGLSTPAASHI  
MRYBEKCTGSIILLASHNPGSPENDMGIKYNLSGGAAPSVTNALWEIKSLTSYK  
IKIIFPELDLGTIGNKKYGPLVDIIDITADYVFLFEIDFDMIKFIYNOSTKN  
WKLLEDSNMGVGTGPGKAI PVDEFGLPADDEVLOQVHPHSPGMPDNLITASSLVK  
RVDEKIRFEGASDGDGRNMIVGYSGVSGSDVAIIAEYAEIPIYFAGOGITGLA  
RSPFETSGAIDRVAKAHGLNCYEVPGKFPALPDAKLSITGESPESTGSNNHREKQ  
GVAMIMAWNLITATYNNKHPENASIKTIQNSFPAKYRTIFTRIDEPKELRANKI  
VDQRAYVTSGVNSAPPADESLKVTQCGDSYTDLGDSVHQGLVYKLSNGARV  
LRLSGTSSGATIRLYIEKYCDKQYQKTAEEYKLINSVYKFLNFKQVLTGEPT  
VRT"  
complement(6411..6455)  
/gene="PGM2"  
/note="PS00710 Phosphoglucosyltransferase and phosphomannomutase  
phosphoserine signature"  
complement(7486..9375)  
/note="YM9718.05c, unknown, len: 629, CAI: 0.13"  
/codon\_start=1  
/product="unknown"  
/protein\_id="CA89742.1"  
/db\_xref="gi:817864"  
/db\_xref="SWISS-PROT:Q04437"

/translation="MSSESTFTVDVSPSMKNNNSKSNALVETLLNKSQRSKTD  
WISCIYLANCVSSENSOEIPNVFQIOSFLAVTTATIGFIRLKQYCDQSHSBSNG  
LOSMIOCLLVASDIKQOFOARKILKQIVFTNLTDLDTIDEIDLLTSSIRILL  
IDGGCQIOERKSKSNMLKLYEALPNRSIYMMNELLVITSPATSVVAFVSGSELRL  
GADISQTSNPGSMODENCICLYEAPASGLNKTAKTAVEVDSQKREYVGV  
KSITTEYIHNENKQVSEDDQSGSITPTTISKDSTAKYRGADVVYLPSTVDTOT  
VYESFPGLDLRGFLNREALRFTLTSSTSTADTRIGCCSDMLAFSALVDVLENK  
IAVARVYSKDSEVNMCAICPVLIEHNNISEKKFVSLTCLRPLFADERVDPEPL  
LDRTTTSGVPLKKKETDQIDELMEQVDSMDTDLPEIPLGNYQPIGEVTDPTLLP  
LPSLNTQOEENKQDPLRIPTVFYRQOQVLEWHOLMINDSREFEIPELPDLSKNT  
SPYTHKFFDSTKLVEVLGIKKVDKLKDSLSLKTLEBEKIPDLETTLLKRGQHSRGP  
MNSNN"  
complement(9806..9878)  
/product="tRNA-Lys"  
/note="tRNA-Lys gene, anticodon, ctc, len: 73"  
complement(10399..10730)  
/note="tRNA-Lys gene, anticodon, ctc, len: 73"  
/note="solo delta element"  
10756..11092  
/note="solo delta element"  
12199..12546  
/note="YM9718.06, unknown, len: 115, CAI: 0.12"  
/codon\_start=1  
/product="unknown"  
/protein\_id="CA89743.1"  
/db\_xref="gi:817865"  
/translation="MGSFMDAFAYDKKADPSVYGNHNTGDSKTQVMSKERYQ  
PRTHQCNLOSMSRSTISGQSDSVDEYVKGRLPAVEIPIKQVNDISMSQGEPLRLVE  
SLRGEFNDKNYNR"  
13269..15332  
/gene="ILV2"  
13269..15332  
/note="YM9718.07, ILV2 gene (SWR1) SW:ILV2 YEAST P07342  
acetylacetate synthase; PS00187 Thiamine pyrophosphate  
enzymes signature"  
/codon\_start=1  
/product="ilv2p"  
/protein\_id="CA89744.1"  
/db\_xref="gi:817866"  
/translation="WIKOSTUKNFAIRKRCROHIAVYNTPMKRSVALAKRPYSSSRYI  
SASPLPASKREPAFSFNVDPLEQPPAPSKLAKLRAPMDTISFVGLTGQITFENMM  
SRQNVDPVFGPGGAILIPVDAIINSKPFNVLPRKHGAGHMAEGVARGSGGVYL  
VTSQGANVTTPMADAFADGIPMVVFTGQVPTAIGTDAFOEADVVIGISSTCKMY  
MYKSEVRLPRLINAEFLATISGRGPVLVDPKVTATILNPIPTKTLPSNLNOL  
TSRQDERFWOSINKADULINKAPVLYNGAGLHNADGRILKEISDRQIYVTTT  
LQGLGSFDQEDPKSLDMLGMHGCATNALAVONADLIYAVGARFDRVTGNISKRAPEA  
RRAAEGRGGIIFHEVSPKINIKVVOQLAVEGADTNLGMMSKIFPVEKRSBMPAO  
INKWKEYPYAVMEETGSGKIPQTVIKKLSKVANDGRHIVTGTGQHQHMAOHV  
TWBPHPHFTSGGLTGWGYGLPAIIGQVAKESLVIDIDGASFNMTLITELSAVOA  
CGPYKTLIINNEBGMVTOHQSLFYEHRYSHTHOLNPDFTKLAAEMGLKGLRVKQEE  
LDAKLKEFVSTKGGVLLVEVDKVPVLPVWAGSSGLDEFINPDEVERQOTELRHKR  
TGGRH"  
14865..14924  
/gene="ILV2"  
/note="PS00187 Thiamine pyrophosphate enzymes signature"  
15772..19431  
/note="YM9718.08, similar to myosin, len: 1219, CAI: 0.20,  
similar to SW:YK09 YEAST P36006 myosin heavy chain  
homologue YKL129C (75.9% identity in 1090 aa overlap);  
PS00017 Arp/GTP-binding site motif A; PS00176 Eukaryotic  
DNA topoisomerase I active site"  
/codon\_start=1  
/product="unknown"  
/protein\_id="CA89745.1"  
/db\_xref="gi:817867"  
/db\_xref="SWISS-PROT:Q04439"

Alignment Scores:  
Pred. No.: 3.73e+03 Length: 35811  
Score: 48.50 Matches: 10  
Percent Similarity: 66.67% Conservative: 2



Best Local Similarity: 55.56% Mismatches: 3  
Query Match: 53.89% Indels: 3  
DB: 8 Gaps: 1

US-10-020-441-3 (1-16) x SC9718 (1-35811)

Search completed: April 8, 2003, 17:14:58  
Job time : 551.771 secs

QY 2 G|YTYG|YArg-----AspAspAsnAspArgAspProSerArgLysAsn 16  
Db 22938 GCGTACGGCTCGTGGACGATGATGATTCAGATAGAGATTCTGAACGACGAAT 22885

RESULT 40

AF510340 886 bp mRNA linear INV 01-JUL-2002  
LOCUS Schistosoma mansoni elastase 2b (CE-2b) mRNA, complete cds.

DEFINITION AF510340  
ACCESSION AF510340  
VERSION AF510340.1 GI:21217532

KEYWORDS Schistosoma mansoni.  
SOURCE Schistosoma mansoni

ORGANISM

REFERENCE

AUTHORS

Salter, J.P., Choe, Y., Albrecht, H., Franklin, C., Lim, K.C.,  
Craig, C.S., and McKerrow, J.H.  
Cercarial elastase is encoded by a functionally conserved gene  
family across multiple species of schistosomes

J. Biol. Chem. 277 (27), 24618-24624 (2002)

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

FEATURES

source

gene

CDS

polya\_signal

BASE COUNT 257 a 165 c 191 g 273 t

ORIGIN

Alignment Scores:  
Pred. No.: 167 Length: 886  
Score: 48.00 Matches: 9  
Percent Similarity: 76.92% Conservative: 1  
Best Local Similarity: 69.23% Mismatches: 3  
Query Match: 53.33% Indels: 0  
DB: 3 Gaps: 0

US-10-020-441-3 (1-16) x AF510340 (1-886)

QY 1 ValG|YTYG|YArgAspAspAsnAspArgAspProSer 13  
Db 484 GTAGGTATGACAGATATAGGATCCTGATCATCC 522

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.4 p5 4578  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 9, 2003, 13:07:49 ; Search time 13.9016 Seconds  
(without alignments)  
237.149 Million cell updates/sec

Title: US-10-020-441-3

Perfect score: 16  
Sequence: 1 VGYGRDNDNDPRSRKN 16

Scoring table: OLIGO  
Gapop 60.0 , Gapext 60.0

Searched: 671580 seqs, 206047115 residues

Word size : 0

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Listing first 1000 summaries

Database :

1: SPREMBL\_21:\*  
2: sp\_archaea:\*  
3: sp\_bacteria:\*  
4: sp\_fungi:\*  
5: sp\_invertebrate:\*  
6: sp\_mammal:\*  
7: sp\_mhc:\*  
8: sp\_organelle:\*  
9: sp\_plant:\*  
10: sp\_rodent:\*  
11: sp\_virus:\*  
12: sp\_vertebrate:\*  
13: sp\_unclassified:\*  
14: sp\_virus:\*  
15: sp\_bacteriap:\*  
16: sp\_archaeap:\*  
17: sp\_archaeap:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	16	100.0	274	5	Q26552
2	14	87.5	257	5	Q26553
3	7	43.8	686	16	Q9A6Y0
4	6	37.5	142	13	Q919J2
5	6	37.5	143	13	Q919J3
6	6	37.5	155	16	Q98N06
7	6	37.5	160	16	Q8Z6S4
8	6	37.5	168	10	Q9ZWM3
9	6	37.5	243	17	P95893
10	6	37.5	244	16	Q8U9E7
11	6	37.5	294	16	Q98HJ8
12	6	37.5	303	16	Q8U9E2
13	6	37.5	308	11	Q8VG21
14	6	37.5	310	11	Q8VFP4
15	6	37.5	335	16	Q8Y618
16					Q8Y618

17	366	2	Q9KK91	Q9KK91 vibrio furn
18	401	13	Q9PRG7	Q9PRG7 xenopus lae
19	412	16	Q53655	Q53655 mycobacteri
20	461	11	Q9DCV1	Q9DCV1 mus musculu
21	478	2	Q9Z190	Q9Z190 lactobacill
22	501	8	Q9GG29	Q9GG29 vallisneria
23	509	16	P72957	P72957 synochocyst
24	515	5	Q8TIN3	Q8TIN3 dictyostell
25	774	5	Q8T4G4	Q8T4G4 drosophila
26	946	10	Q9C5J3	Q9C5J3 arabidopsis
27	985	5	Q9Y142	Q9Y142 drosophila
28	985	5	Q9V4X8	Q9V4X8 drosophila
29	1955	10	Q8W078	Q8W078 oryza sativ
30	4215	5	Q9W332	Q9W332 drosophila
31	40	12	Q91WZ3	Q91WZ3 human echov
32	40	12	Q91WY8	Q91WY8 human coxa
33	50	12	Q91TV4	Q91TV4 human echov
34	51	12	Q66826	Q66826 human echov
35	63	16	Q8X2J9	Q8X2J9 escherichia
36	75	9	Q21995	Q21995 streptococc
37	75	16	Q9K1L9	Q9K1L9 neisseria m
38	76	16	Q9WY01	Q9WY01 thermotoga
39	80	16	Q8XBR5	Q8XBR5 escherichia
40	86	2	Q8VR70	Q8VR70 escherichia
41	88	11	Q8VCU9	Q8VCU9 mus musculu
42	89	17	Q8TVQ8	Q8TVQ8 methanobarc
43	90	2	Q8VU92	Q8VU92 lactobacill
44	96	2	Q07160	Q07160 mycobacteri
45	97	10	Q86BE4	Q86BE4 oryza sativ
46	107	3	Q9Y813	Q9Y813 echizosacch
47	109	5	Q21398	Q21398 caenorhabdi
48	110	16	Q92V18	Q92V18 rhizobium m
49	111	10	Q9FIB1	Q9FIB1 arabidopsis
50	111	11	Q9DAP8	Q9DAP8 mus musculu
51	121	16	Q8VCK6	Q8VCK6 brucella me
52	122	5	Q25669	Q25669 plasmodium
53	122	5	Q25670	Q25670 plasmodium
54	122	5	Q25671	Q25671 plasmodium
55	125	17	Q29944	Q29944 archaeoglob
56	127	16	Q9A870	Q9A870 caulobacter
57	128	16	P74747	P74747 synochocyst
58	128	16	Q984X7	Q984X7 rhizobium l
59	128	17	Q9V2F8	Q9V2F8 pyrococcus
60	129	13	Q90514	Q90514 fuqu rubrip
61	131	5	Q9NDH9	Q9NDH9 plasmodium
62	131	10	Q9C9G0	Q9C9G0 arabidopsis
63	133	12	Q912H3	Q912H3 human coxa
64	134	16	Q8Y0C2	Q8Y0C2 brucella me
65	135	4	Q9UGW6	Q9UGW6 homo sapien
66	137	12	Q9U940	Q9U940 human coxa
67	137	12	Q9U939	Q9U939 human coxa
68	137	12	Q9U938	Q9U938 human coxa
69	137	12	Q9U937	Q9U937 human coxa
70	137	12	Q9U936	Q9U936 human coxa
71	137	12	Q9U935	Q9U935 human coxa
72	137	12	Q9U934	Q9U934 human coxa
73	137	12	Q9U933	Q9U933 human coxa
74	137	12	Q9U932	Q9U932 human coxa
75	137	12	Q9U930	Q9U930 human coxa
76	137	12	Q9U929	Q9U929 human coxa
77	137	12	Q9U928	Q9U928 human coxa
78	137	12	Q9U927	Q9U927 human coxa
79	137	12	Q9U926	Q9U926 human coxa
80	137	12	Q9U925	Q9U925 human coxa
81	137	12	Q9U924	Q9U924 human coxa
82	137	12	Q9U923	Q9U923 human coxa
83	137	12	Q9U922	Q9U922 human coxa
84	137	12	Q9U921	Q9U921 human coxa
85	137	12	Q9U920	Q9U920 human coxa
86	137	12	Q9U919	Q9U919 human coxa
87	137	12	Q9U918	Q9U918 human coxa
88	137	12	Q9U917	Q9U917 human coxa
89	137	12	Q9U916	Q9U916 human coxa

90	5	31.2	137	12	Q9J915	Q9J915 human coxa	163	5	31.2	224	5	Q8T5V0	Q8T5V0 anophelae s
91	5	31.2	137	12	Q9J914	Q9J914 human coxa	164	5	31.2	224	16	Q9WZB1	Q9WZB1 thermotoga
92	5	31.2	137	12	Q9J913	Q9J913 human coxa	165	5	31.2	225	5	Q9S5B3	Q9S5B3 drosophila
93	5	31.2	137	12	Q9J912	Q9J912 human coxa	166	5	31.2	226	16	Q8X8N5	Q8X8N5 ralatonia s
94	5	31.2	137	12	Q9J911	Q9J911 human coxa	167	5	31.2	227	10	Q9S5R3	Q9S5R3 arbidopsis
95	5	31.2	137	12	Q9J910	Q9J910 human coxa	168	5	31.2	229	5	Q20490	Q20490 caenorhabdi
96	5	31.2	137	12	Q9J909	Q9J909 human coxa	169	5	31.2	231	10	Q949A9	Q949A9 oryza sativ
97	5	31.2	137	12	Q9J908	Q9J908 human coxa	170	5	31.2	232	5	P81223	P81223 glardia lam
98	5	31.2	137	12	Q9J907	Q9J907 human coxa	171	5	31.2	235	16	Q8YH07	Q8YH07 bruceella me
99	5	31.2	138	16	Q98HD9	Q98HD9 rhizobium 1	172	5	31.2	241	16	Q8HA16	Q8HA16 rhizobium 1
100	5	31.2	139	10	Q9L7C5	Q9L7C5 arbidopsis	173	5	31.2	242	2	Q50391	Q50391 mycobacteri
101	5	31.2	141	10	Q8Z4V0	Q8Z4V0 salmoneilla	174	5	31.2	242	10	Q387J3	Q387J3 antirrhinum
102	5	31.2	143	16	Q8RUE1	Q8RUE1 oryza sativ	175	5	31.2	243	2	Q9AGK5	Q9AGK5 streptococ
103	5	31.2	146	12	Q91QW4	Q91QW4 coxsackielevi	176	5	31.2	243	10	Q9C7H1	Q9C7H1 arbidopsis
104	5	31.2	146	12	Q91QV1	Q91QV1 human echov	177	5	31.2	244	16	Q9A858	Q9A858 caulobacter
105	5	31.2	146	12	Q91QW5	Q91QW5 human echov	178	5	31.2	245	9	Q94M62	Q94M62 streptococ
106	5	31.2	146	12	Q91QW5	Q91QW5 human coxa	179	5	31.2	248	16	Q8XPNT	Q8XPNT ralatonia s
107	5	31.2	150	15	Q8UTV2	Q8UTV2 human immun	180	5	31.2	249	4	Q9NUTC5	Q9NUTC5 homo sapien
108	5	31.2	151	16	Q97KP1	Q97KP1 clostridium	181	5	31.2	251	10	Q9LEP3	Q9LEP3 betula verr
109	5	31.2	152	16	Q97SW6	Q97SW6 streptococ	182	5	31.2	252	5	Q9W5S0	Q9W5S0 drosophila
110	5	31.2	158	16	Q9ZFN1	Q9ZFN1 listeria in	183	5	31.2	255	10	Q9FU00	Q9FU00 oryza sativ
111	5	31.2	160	16	Q8YSU9	Q8YSU9 anabena sp	184	5	31.2	256	11	Q924A9	Q924A9 mus musculu
112	5	31.2	160	16	Q05882	Q05882 mycobacteri	185	5	31.2	256	11	Q91W41	Q91W41 mus musculu
113	5	31.2	163	3	Q13352	Q13352 saccharomyc	186	5	31.2	256	12	Q9UTD8	Q9UTD8 cucurbit le
114	5	31.2	163	3	Q8X9Z5	Q8X9Z5 escherichia	187	5	31.2	256	12	Q8QZ70	Q8QZ70 cucurbit le
115	5	31.2	165	16	Q07535	Q07535 bacillus su	188	5	31.2	258	16	Q92U84	Q92U84 rhizobium m
116	5	31.2	165	16	Q8RAN4	Q8RAN4 thermoaer	189	5	31.2	261	4	Q9UTJ3	Q9UTJ3 homo sapien
117	5	31.2	167	2	Q45732	Q45732 bacillus th	190	5	31.2	265	2	Q93WJ6	Q93WJ6 corynebacte
118	5	31.2	167	2	Q45741	Q45741 bacillus th	191	5	31.2	265	10	Q80699	Q80699 arbidopsis
119	5	31.2	167	10	Q8W5F2	Q8W5F2 oryza sativ	192	5	31.2	268	10	Q9FMT7	Q9FMT7 arbidopsis
120	5	31.2	167	16	Q92G72	Q92G72 rickettsia	193	5	31.2	270	4	Q8TD38	Q8TD38 homo sapien
121	5	31.2	168	10	Q9AS06	Q9AS06 oryza sativ	194	5	31.2	270	16	Q98NM3	Q98NM3 rhizobium 1
122	5	31.2	168	10	Q9FHP6	Q9FHP6 arbidopsis	195	5	31.2	270	17	Q8TVI0	Q8TVI0 methanopyru
123	5	31.2	170	5	Q95U88	Q95U88 helicoverpa	196	5	31.2	272	4	Q8TDJ7	Q8TDJ7 homo sapien
124	5	31.2	171	10	Q9C8B5	Q9C8B5 arbidopsis	197	5	31.2	272	16	Q985D6	Q985D6 rhizobium 1
125	5	31.2	172	3	Q9P922	Q9P922 pneumocysti	198	5	31.2	275	2	Q87248	Q87248 lactococcus
126	5	31.2	174	2	Q06444	Q06444 bacillus st	199	5	31.2	275	2	Q85289	Q85289 sphingomona
127	5	31.2	174	16	Q9L0U8	Q9L0U8 streptomyce	200	5	31.2	276	2	Q69127	Q69127 burkholderi
128	5	31.2	176	2	Q9RCU0	Q9RCU0 streptococ	201	5	31.2	276	2	Q9AEBU9	Q9AEBU9 burkholderi
129	5	31.2	178	2	Q9RAP7	Q9RAP7 leptospira	202	5	31.2	278	4	Q9BVH0	Q9BVH0 homo sapien
130	5	31.2	178	16	Q9A8R6	Q9A8R6 caulobacter	203	5	31.2	280	10	Q94JX8	Q94JX8 oryza sativ
131	5	31.2	178	16	Q8ZNB4	Q8ZNB4 salmoneilla	204	5	31.2	285	7	Q31521	Q31521 poecilia re
132	5	31.2	182	12	Q91FG9	Q91FG9 chilo iride	205	5	31.2	286	4	Q96MH2	Q96MH2 homo sapien
133	5	31.2	183	5	Q9W105	Q9W105 drosophila	206	5	31.2	287	10	Q82503	Q82503 arbidopsis
134	5	31.2	187	16	Q8UDS9	Q8UDS9 agrobacteri	207	5	31.2	287	11	Q9DCQ2	Q9DCQ2 mus musculu
135	5	31.2	191	17	Q9HRX6	Q9HRX6 agrobacteri	208	5	31.2	287	11	Q9VCC0	Q9VCC0 mus musculu
136	5	31.2	192	16	Q9F3L6	Q9F3L6 streptomyce	209	5	31.2	289	11	Q9R0Y0	Q9R0Y0 mus musculu
137	5	31.2	198	2	Q9RLI7	Q9RLI7 neisseria m	210	5	31.2	290	11	Q91UZ5	Q91UZ5 mus musculu
138	5	31.2	198	2	Q9RLI8	Q9RLI8 neisseria m	211	5	31.2	291	3	Q9Y7C5	Q9Y7C5 emericella
139	5	31.2	198	2	Q9RLI9	Q9RLI9 neisseria m	212	5	31.2	292	11	Q925K1	Q925K1 mus musculu
140	5	31.2	198	2	Q9RLM0	Q9RLM0 neisseria m	213	5	31.2	294	2	Q9XCJ4	Q9XCJ4 bartonella
141	5	31.2	198	16	Q9FBK0	Q9FBK0 streptomyce	214	5	31.2	296	3	Q9P4U9	Q9P4U9 alternaria
142	5	31.2	200	9	Q8SCV3	Q8SCV3 pseudomona	215	5	31.2	296	3	Q9CIC4	Q9CIC4 ajellomyces
143	5	31.2	200	16	Q34911	Q34911 bacillus su	216	5	31.2	297	4	Q8WUE1	Q8WUE1 homo sapien
144	5	31.2	204	16	Q8VKI2	Q8VKI2 mycobacteri	217	5	31.2	297	4	Q9NP96	Q9NP96 homo sapien
145	5	31.2	204	16	Q98HNO	Q98HNO rhizobium 1	218	5	31.2	297	11	Q9CRD2	Q9CRD2 mus musculu
146	5	31.2	207	17	Q27672	Q27672 methanobact	219	5	31.2	298	11	Q35726	Q35726 marmota max
147	5	31.2	209	5	Q21621	Q21621 caenorhabdi	220	5	31.2	299	4	Q15560	Q15560 homo sapien
148	5	31.2	210	2	Q9EXB4	Q9EXB4 neisseria m	221	5	31.2	305	17	Q8TUS6	Q8TUS6 methanocarc
149	5	31.2	210	4	Q9BSX1	Q9BSX1 homo sapien	222	5	31.2	306	16	Q9WU08	Q9WU08 staphylococ
150	5	31.2	211	2	Q9AOD5	Q9AOD5 lactococcus	223	5	31.2	307	16	Q8UD00	Q8UD00 agrobacteri
151	5	31.2	212	2	P94508	P94508 bacillus su	224	5	31.2	308	9	Q8SCH8	Q8SCH8 propionibac
152	5	31.2	215	4	Q9UFU4	Q9UFU4 homo sapien	225	5	31.2	309	10	Q8SE63	Q8SE63 oryza sativ
153	5	31.2	215	5	Q9VSR9	Q9VSR9 drosophila	226	5	31.2	309	16	Q9XK01	Q9XK01 rhizobium m
154	5	31.2	216	4	Q96FZ5	Q96FZ5 homo sapien	227	5	31.2	309	16	Q92UG2	Q92UG2 rhizobium m
155	5	31.2	218	5	Q9XYL9	Q9XYL9 fasciola gi	228	5	31.2	311	10	Q39500	Q39500 cylindrothe
156	5	31.2	220	16	Q9L1B5	Q9L1B5 streptomyce	229	5	31.2	311	13	Q99162	Q99162 xiphophorus
157	5	31.2	221	15	Q73155	Q73155 human immun	230	5	31.2	311	16	Q8ZAU9	Q8ZAU9 yerstinia pe
158	5	31.2	221	16	Q988S5	Q988S5 rhizobium 1	231	5	31.2	314	12	Q9YOX5	Q9YOX5 porcine cir
159	5	31.2	222	5	Q9XXI7	Q9XXI7 caenorhabdi	232	5	31.2	314	17	Q8TWH3	Q8TWH3 methanopyru
160	5	31.2	223	5	Q8SYV2	Q8SYV2 drosophila	233	5	31.2	315	4	Q9UUT4	Q9UUT4 homo sapien
161	5	31.2	223	10	Q8S2A1	Q8S2A1 oryza sativ	234	5	31.2	317	5	Q26564	Q26564 echistosoma
162	5	31.2	223	16	Q92Y23	Q92Y23 rhizobium m	235	5	31.2	317	10	Q9L039	Q9L039 arbidopsis

236	5	31.2	320	2	Q9R6V8	Q9r6v8	bynechococc	309	5	31.2	382	2	Q93L43	Q93l43	bacillus we
237	5	31.2	320	10	Q9M9D9	Q9m9d9	arabidopsis	310	5	31.2	382	2	Q93L42	Q93l42	Q93l42 bacillus we
238	5	31.2	322	3	Q9URQ3	Q9urq3	saccharomyc	311	5	31.2	382	3	Q8WZV1	Q8wzv1	neurospora
239	5	31.2	322	12	Q4O9Z0	Q4o9z0	kaposi's sa	312	5	31.2	384	16	Q9KX59	Q9kx59	streptomyces
240	5	31.2	323	17	Q8TQZ6	Q8tqz6	methanopyru	313	5	31.2	385	5	Q179G9	Q179g9	caenorhabdi
241	5	31.2	324	2	Q064E8	Q064e8	neisseria m	314	5	31.2	386	2	Q87874	Q87874	thauera aro
242	5	31.2	324	17	Q282Z0	Q282z0	archaeoglob	315	5	31.2	387	3	P87213	P87213	sclerotinia
243	5	31.2	326	16	Q9CFZ2	Q9cfz2	lactococcus	316	5	31.2	388	11	Q91VMS	Q91vms	mus muscu
244	5	31.2	328	15	Q9ER18	Q9er18	human immun	317	5	31.2	389	3	Q9Y775	Q9y775	candida tro
245	5	31.2	328	15	Q9EEN6	Q9een6	human immun	318	5	31.2	389	17	Q8TRYG0	Q8tryg0	methanopyru
246	5	31.2	328	15	Q9ED97	Q9ed97	human immun	319	5	31.2	390	4	Q96E53	Q96e53	homo sapien
247	5	31.2	330	16	Q9RUR8	Q9rur8	deinococcus	320	5	31.2	391	2	Q9RAP3	Q9rap3	leptospira
248	5	31.2	331	4	Q9NPA7	Q9npa7	homo sapien	321	5	31.2	391	4	Q969R3	Q969r3	homo sapien
249	5	31.2	331	5	Q9NLO9	Q9nlo9	branchiosto	322	5	31.2	392	11	Q9WV02	Q9wv02	mus muscu
250	5	31.2	331	5	Q108J4	Q108j4	schistosoma	323	5	31.2	392	5	Q9NEV3	Q9nev3	caenorhabdi
251	5	31.2	333	12	P90495	P90495	kaposi's sa	324	5	31.2	392	13	Q9PTE7	Q9pte7	ambystoma t
252	5	31.2	335	10	Q9M034	Q9m034	arabidopsis	325	5	31.2	394	2	Q9RCA3	Q9rcA3	agrobacteri
253	5	31.2	335	10	Q8RZC0	Q8rzc0	oryza sativ	326	5	31.2	394	10	Q9LRF04	Q9lrf04	arabidopsis
254	5	31.2	336	4	Q9BVH7	Q9bvH7	h simlilar t	327	5	31.2	394	16	Q9PR19	Q9pr19	ureaplasma
255	5	31.2	336	10	Q9FZ81	Q9fz81	arabidopsis	328	5	31.2	395	4	Q8TDU4	Q8tdU4	homo sapien
256	5	31.2	336	16	Q8ZHN1	Q8zhn1	yerania pe	329	5	31.2	396	10	Q43140	Q43140	sesamum ind
257	5	31.2	339	10	Q9FPE7	Q9ffe7	arabidopsis	330	5	31.2	396	10	Q9M880	Q9m880	arabidopsis
258	5	31.2	340	17	Q9UTD9	Q9utd9	pyrococcus	331	5	31.2	396	10	Q65040	Q65040	macadyena
259	5	31.2	341	2	Q9AQ02	Q9aq02	bradyrhizob	332	5	31.2	396	10	Q43141	Q43141	sesamum ind
260	5	31.2	344	10	Q9SQH2	Q9sqh2	brassica na	333	5	31.2	396	10	Q24498	Q24498	helianthus
261	5	31.2	346	13	Q9Y4J5	Q9y4j5	homo sapien	334	5	31.2	396	10	Q93XZ0	Q93xz0	sesamum ind
262	5	31.2	346	17	P17776	P17776	xiphophorus	335	5	31.2	396	10	Q8VXJ8	Q8vxj8	helianthus
263	5	31.2	346	17	Q8ZYL5	Q8zyl5	pyrobaculum	336	5	31.2	396	10	Q8VXJ7	Q8vxj7	helianthus
264	5	31.2	348	17	Q97BD3	Q97bd3	thermoplasma	337	5	31.2	398	16	Q930R4	Q930r4	rhizobium m
265	5	31.2	349	10	Q9FEA9	Q9fea9	oryza sativ	338	5	31.2	398	16	Q927Y8	Q927y8	liesteria in
266	5	31.2	351	16	Q8RI88	Q8ri88	fusobacteri	339	5	31.2	399	13	Q9W7R1	Q9w7r1	oryzae lat
267	5	31.2	352	5	Q9VAR3	Q9vaE3	drosophila	340	5	31.2	400	11	Q70220	Q70220	mus muscu
268	5	31.2	353	17	Q975X9	Q975x9	sulfolobus	341	5	31.2	400	11	Q9JLNT7	Q9jlnT7	rattus rat
269	5	31.2	358	7	Q62884	Q62884	ictalurus p	342	5	31.2	400	11	Q9JLJ18	Q9jlj18	mus muscu
270	5	31.2	358	7	Q62895	Q62895	ictalurus p	343	5	31.2	400	16	Q9HWM2	Q9hwm2	pseudomonas
271	5	31.2	359	4	Q96EN2	Q96en2	homo sapien	344	5	31.2	401	8	Q9XQR7	Q9xqr7	brassica ju
272	5	31.2	359	5	Q8TOR7	Q8tor7	drosophila	345	5	31.2	401	16	Q9RE04	Q9re04	liesteria mo
273	5	31.2	359	7	Q98192	Q98192	acrocephalu	346	5	31.2	401	16	Q8XWL3	Q8xwl3	ralstonia s
274	5	31.2	360	7	Q98191	Q98191	acrocephalu	347	5	31.2	402	4	Q9NS06	Q9ns06	homo sapien
275	5	31.2	361	7	Q62896	Q62896	ictalurus p	348	5	31.2	402	5	Q8N0S2	Q8n0s2	giardia lam
276	5	31.2	362	11	Q9R219	Q9r219	cricetulus	349	5	31.2	403	4	Q9C009	Q9c009	homo sapien
277	5	31.2	362	11	Q9R220	Q9r220	cricetulus	350	5	31.2	406	3	Q9P8D1	Q9p8d1	cephalospo
278	5	31.2	362	11	Q35886	Q35886	cricetulus	351	5	31.2	406	10	Q8VY21	Q8vy21	arabidopsis
279	5	31.2	365	10	Q04580	Q04580	arabidopsis	352	5	31.2	406	16	Q9RRP8	Q9rrp8	deinococcus
280	5	31.2	366	10	Q9FZD6	Q9fzD6	arabidopsis	353	5	31.2	407	17	Q9V107	Q9v107	pyrococcus
281	5	31.2	366	17	Q97Y10	Q97y10	sulfolobus	354	5	31.2	407	10	Q82257	Q82257	arabidopsis
282	5	31.2	367	5	Q9SS40	Q9ssa0	drosophila	355	5	31.2	407	17	Q8U0N1	Q8u0n1	pyrococcus
283	5	31.2	367	5	Q9VIP8	Q9vip8	drosophila	356	5	31.2	408	2	Q05593	Q05593	actinomyces
284	5	31.2	367	5	Q8TJF8	Q8tjF8	caenorhabdi	357	5	31.2	409	10	Q9SM87	Q9sm87	aracheis hyp
285	5	31.2	367	7	Q9GIW7	Q9giw7	ictalurus p	358	5	31.2	409	17	Q59207	Q59207	pyrococcus
286	5	31.2	367	16	Q8RC00	Q8rc00	thermoanaer	359	5	31.2	414	2	P70719	P70719	actinobacil
287	5	31.2	368	10	Q93WR8	Q93wr8	medicago va	360	5	31.2	414	11	Q9DAU0	Q9daU0	mus muscu
288	5	31.2	370	2	Q06454	Q06454	neisseria m	361	5	31.2	414	13	Q90Z99	Q90z99	xenopus lae
289	5	31.2	370	2	Q06467	Q06467	neisseria g	362	5	31.2	414	17	Q97AF6	Q97af6	thermoplasma
290	5	31.2	370	3	Q9P358	Q9p358	aspergillus	363	5	31.2	420	10	Q22345	Q22345	zea mays (m
291	5	31.2	370	3	Q9P4W4	Q9p4w4	aspergillus	364	5	31.2	420	13	Q91184	Q91184	oryzae lat
292	5	31.2	370	5	Q8SUY4	Q8suy4	encephalito	365	5	31.2	420	16	Q9RKC5	Q9rkC5	streptomyces
293	5	31.2	371	16	Q8X6U6	Q8x6U6	escherichia	366	5	31.2	427	2	Q9RG72	Q9rg72	coxiella bu
294	5	31.2	372	16	Q981E7	Q981e7	rhizobium l	367	5	31.2	427	2	Q9RG71	Q9rg71	coxiella bu
295	5	31.2	375	10	Q9C7G4	Q9c7g4	campylobact	368	5	31.2	427	2	Q9RG70	Q9rg70	coxiella bu
296	5	31.2	376	2	P71123	P71123	campylobact	369	5	31.2	427	2	Q9RG69	Q9rg69	coxiella bu
297	5	31.2	377	7	Q9GIW6	Q9giw6	ictalurus p	370	5	31.2	427	2	Q9RG68	Q9rg68	coxiella bu
298	5	31.2	378	5	Q17183	Q17183	caenorhabdi	371	5	31.2	427	2	Q9RG67	Q9rg67	coxiella bu
299	5	31.2	378	13	Q916B3	Q916b3	xenopus lae	372	5	31.2	427	2	Q9RG66	Q9rg66	coxiella bu
300	5	31.2	379	3	Q9CA41	Q9ca41	alternaria	373	5	31.2	427	2	Q9RG65	Q9rg65	coxiella bu
301	5	31.2	379	3	Q9CA40	Q9ca40	alternaria	374	5	31.2	427	2	Q9RG64	Q9rg64	coxiella bu
302	5	31.2	379	12	Q9Y1E5	Q9y1e5	lymantria d	375	5	31.2	427	2	Q9RG63	Q9rg63	coxiella bu
303	5	31.2	380	3	Q9Y7W1	Q9y7w1	botrytis ci	376	5	31.2	427	2	Q9RG62	Q9rg62	coxiella bu
304	5	31.2	380	10	Q8W5I5	Q8w5i5	oryza sativ	377	5	31.2	427	2	Q9ZH99	Q9zh99	coxiella bu
305	5	31.2	380	16	Q9CHN1	Q9chn1	lactococcus	378	5	31.2	427	17	Q8TP83	Q8tp83	methanobact
306	5	31.2	381	10	Q9FRV5	Q9frv5	arabidopsis	379	5	31.2	428	5	Q9GRP1	Q9grp1	leishmania
307	5	31.2	381	12	Q8QGA1	Q8qga1	felis domes	380	5	31.2	429	4	Q43729	Q43729	homo sapien
308	5	31.2	382	2	Q93L44	Q93l44	bacillus we	381	5	31.2	430	16	Q9ZQL6	Q9zql6	rhizobium m

382	5	31.2	434	16	Q917A9	Q917A9 pseudomonas	455	5	31.2	525	3	Q9P6W8	Q9P6W8 neurospora
383	5	31.2	435	16	Q8UD77	Q8UD77 agrobacteri	456	5	31.2	525	12	Q8J724	Q8J724 measles vir
384	5	31.2	436	10	Q9FLK0	Q9FLK0 aradidopsis	457	5	31.2	529	2	Q9PD13	Q9PD13 breviacter
385	5	31.2	436	17	Q27786	Q27786 methanobact	458	5	31.2	530	5	Q9VHW0	Q9VHW0 drosophila
386	5	31.2	438	16	Q53379	Q53379 mycobacteri	459	5	31.2	533	2	Q9F1X4	Q9F1X4 rhodobacter
387	5	31.2	439	10	Q9LY67	Q9LY67 aradidopsis	460	5	31.2	534	2	Q68165	Q68165 streptococc
388	5	31.2	443	5	Q9V8G4	Q9V8G4 drosophila	461	5	31.2	534	2	Q33898	Q33898 streptococc
389	5	31.2	443	16	Q8XZM7	Q8XZM7 ralestoma s	462	5	31.2	535	17	Q27987	Q27987 archaeoglob
390	5	31.2	445	16	Q9KD63	Q9KD63 bacillus ha	463	5	31.2	536	16	Q8U6A8	Q8U6A8 agrobacteri
391	5	31.2	446	12	Q9DOX6	Q9DOX6 equine coro	464	5	31.2	536	17	Q96ZU8	Q96ZU8 sulfolobus
392	5	31.2	446	16	Q98A01	Q98A01 rhizobium 1	465	5	31.2	537	10	Q9SHS8	Q9SHS8 aradidopsis
393	5	31.2	447	16	Q92V52	Q92V52 rhizobium m	466	5	31.2	544	10	Q9WDX9	Q9WDX9 aradidopsis
394	5	31.2	448	12	Q9QAR8	Q9QAR8 bovine coro	467	5	31.2	547	5	Q9V4U1	Q9V4U1 drosophila
395	5	31.2	448	12	Q9QAR1	Q9QAR1 bovine coro	468	5	31.2	553	16	Q9RSN4	Q9RSN4 delnoccocus
396	5	31.2	448	12	Q91A22	Q91A22 bovine coro	469	5	31.2	555	2	Q85737	Q85737 streptomyce
397	5	31.2	448	12	Q8VA32	Q8VA32 bovine coro	470	5	31.2	555	15	Q90D61	Q90D61 human immun
398	5	31.2	451	16	Q92I90	Q92I90 rhizobium m	471	5	31.2	555	16	Q8U8S0	Q8U8S0 agrobacteri
399	5	31.2	453	2	Q9AH00	Q9AH00 pseudomonas	472	5	31.2	556	10	Q9F282	Q9F282 aradidopsis
400	5	31.2	454	16	Q67843	Q67843 aquifex aeo	473	5	31.2	558	16	Q8XD37	Q8XD37 escherichia
401	5	31.2	455	10	Q9SC69	Q9SC69 oryza sativ	474	5	31.2	563	16	Q8ZCA2	Q8ZCA2 yersinia pe
402	5	31.2	456	2	Q9APM5	Q9APM5 bilophila w	475	5	31.2	564	10	Q9SM88	Q9SM88 oryza sativ
403	5	31.2	456	2	Q9N5N7	Q9N5N7 caenorhabdi	476	5	31.2	564	16	Q92TUI	Q92TUI rhizobium m
404	5	31.2	456	16	Q915C2	Q915C2 pseudomonas	477	5	31.2	565	16	Q9RJH0	Q9RJH0 streptomyce
405	5	31.2	457	5	Q18534	Q18534 caenorhabdi	478	5	31.2	569	16	Q9ZBP0	Q9ZBP0 streptomyce
406	5	31.2	458	16	Q98A14	Q98A14 rhizobium 1	479	5	31.2	569	16	Q9RU39	Q9RU39 delnoccocus
407	5	31.2	459	16	Q92WD3	Q92WD3 rhizobium m	480	5	31.2	572	16	Q54207	Q54207 streptomyce
408	5	31.2	459	16	Q98L27	Q98L27 rhizobium 1	481	5	31.2	573	6	Q9TV71	Q9TV71 equus cabal
409	5	31.2	459	16	Q987B2	Q987B2 rhizobium 1	482	5	31.2	573	10	Q81727	Q81727 aradidopsis
410	5	31.2	460	5	Q9V5S1	Q9V5S1 drosophila	483	5	31.2	573	12	Q994E5	Q994E5 porcine ade
411	5	31.2	463	5	Q9BL37	Q9BL37 caenorhabdi	484	5	31.2	573	17	Q9UWM2	Q9UWM2 sulfolobus
412	5	31.2	466	16	Q8VU28	Q8VU28 mycobacteri	485	5	31.2	576	16	Q8R9B5	Q8R9B5 thermoanaer
413	5	31.2	467	16	Q9A3S0	Q9A3S0 caulobacter	486	5	31.2	584	10	Q22346	Q22346 zea mays (m
414	5	31.2	467	16	Q8ZNP9	Q8ZNP9 salmonella	487	5	31.2	584	16	Q92YK4	Q92YK4 rhizobium m
415	5	31.2	467	16	Q825F4	Q825F4 salmonella	488	5	31.2	586	11	Q9DCM7	Q9DCM7 mus musculu
416	5	31.2	469	4	Q14573	Q14573 homo sapien	489	5	31.2	586	11	Q8VDF8	Q8VDF8 mus musculu
417	5	31.2	474	16	Q84286	Q84286 chlamydia t	490	5	31.2	587	4	Q96BF6	Q96BF6 homo sapien
418	5	31.2	475	16	Q8YFE5	Q8YFE5 bruceella me	491	5	31.2	588	5	Q964F6	Q964F6 plasmodium
419	5	31.2	476	3	Q12186	Q12186 saccharomyc	492	5	31.2	588	16	Q9A696	Q9A696 caulobacter
420	5	31.2	476	10	Q92WM0	Q92WM0 oryza sativ	493	5	31.2	589	10	Q9S1Z3	Q9S1Z3 aradidopsis
421	5	31.2	477	4	Q96963	Q96963 homo sapien	494	5	31.2	590	4	Q9H6W1	Q9H6W1 homo sapien
422	5	31.2	478	2	Q30615	Q30615 myxococcus	495	5	31.2	591	16	Q92UW6	Q92UW6 rhizobium m
423	5	31.2	478	5	Q93759	Q93759 caenorhabdi	496	5	31.2	592	4	Q96CH9	Q96CH9 homo sapien
424	5	31.2	479	16	Q9HUD2	Q9HUD2 pseudomonas	497	5	31.2	592	10	Q9SB60	Q9SB60 aradidopsis
425	5	31.2	482	5	Q9VLM3	Q9VLM3 drosophila	498	5	31.2	593	4	Q96NM3	Q96NM3 homo sapien
426	5	31.2	483	5	Q9VLM2	Q9VLM2 drosophila	499	5	31.2	595	5	Q19745	Q19745 caenorhabdi
427	5	31.2	486	5	Q9VLM4	Q9VLM4 drosophila	500	5	31.2	595	11	Q921G3	Q921G3 mus musculu
428	5	31.2	486	16	Q8YMI6	Q8YMI6 anabaena sp	501	5	31.2	595	17	Q28508	Q28508 archaeoglob
429	5	31.2	487	5	Q968U5	Q968U5 trypanosoma	502	5	31.2	596	5	Q964F7	Q964F7 plasmodium
430	5	31.2	487	17	Q8TND1	Q8TND1 methanosarc	503	5	31.2	597	4	Q969V8	Q969V8 homo sapien
431	5	31.2	489	10	Q48888	Q48888 zea mays (m	504	5	31.2	597	11	Q9R0V6	Q9R0V6 mus musculu
432	5	31.2	493	4	Q9NX26	Q9NX26 homo sapien	505	5	31.2	599	16	Q92RD6	Q92RD6 rhizobium m
433	5	31.2	493	4	Q9NUD9	Q9NUD9 homo sapien	506	5	31.2	600	5	Q964F5	Q964F5 plasmodium
434	5	31.2	495	16	Q05896	Q05896 mycobacteri	507	5	31.2	600	5	Q95NL3	Q95NL3 plasmodium
435	5	31.2	496	2	Q9AKI1	Q9AKI1 rickettsia	508	5	31.2	601	5	Q22172	Q22172 caenorhabdi
436	5	31.2	496	2	Q9AKN4	Q9AKN4 rickettsia	509	5	31.2	604	5	Q964F8	Q964F8 plasmodium
437	5	31.2	496	16	Q92J41	Q92J41 rickettsia	510	5	31.2	604	16	Q8UPA5	Q8UPA5 agrobacteri
438	5	31.2	500	2	Q9F3V7	Q9F3V7 pseudonocac	511	5	31.2	608	16	Q9CBU6	Q9CBU6 mycobacteri
439	5	31.2	500	5	Q9U2J0	Q9U2J0 caenorhabdi	512	5	31.2	608	16	Q8U8H1	Q8U8H1 agrobacteri
440	5	31.2	501	2	Q83012	Q83012 bacillus sp	513	5	31.2	609	16	Q8Z9S8	Q8Z9S8 yersinia pe
441	5	31.2	501	16	Q8YH64	Q8YH64 bruceella me	514	5	31.2	611	10	Q9LKB4	Q9LKB4 aradidopsis
442	5	31.2	504	6	Q95JF8	Q95JF8 sus scrofa	515	5	31.2	611	10	Q8YVDR	Q8YVDR mus musculu
443	5	31.2	504	6	Q8SPK1	Q8SPK1 sus scrofa	516	5	31.2	612	16	Q31527	Q31527 bacillus su
444	5	31.2	504	6	Q8SPK0	Q8SPK0 sus scrofa	517	5	31.2	613	11	P81718	P81718 rattus norv
445	5	31.2	507	4	Q9NOH7	Q9NOH7 homo sapien	518	5	31.2	614	5	Q95R23	Q95R23 babesia ros
446	5	31.2	507	4	Q9BV27	Q9BV27 homo sapien	519	5	31.2	622	11	Q8R580	Q8R580 mus musculu
447	5	31.2	515	3	Q60099	Q60099 schizosacch	520	5	31.2	624	4	Q9UK67	Q9UK67 homo sapien
448	5	31.2	516	2	Q8VMG2	Q8VMG2 pseudomonas	521	5	31.2	624	4	Q96AJ8	Q96AJ8 homo sapien
449	5	31.2	516	16	Q9ABD2	Q9ABD2 caulobacter	522	5	31.2	625	10	Q95X21	Q95X21 aradidopsis
450	5	31.2	517	12	Q9YU81	Q9YU81 turkey aden	523	5	31.2	625	11	Q61468	Q61468 mus musculu
451	5	31.2	519	4	Q9NY18	Q9NY18 homo sapien	524	5	31.2	626	2	Q8RPA3	Q8RPA3 methyllobact
452	5	31.2	520	16	Q8UEB6	Q8UEB6 agrobacteri	525	5	31.2	627	16	Q8ZOU9	Q8ZOU9 anabaena sp
453	5	31.2	522	11	Q8R585	Q8R585 mus musculu	526	5	31.2	629	5	Q95R24	Q95R24 babesia ros
454	5	31.2	523	2	Q9WV77	Q9WV77 agrobacteri	527	5	31.2	630	3	Q9P4R5	Q9P4R5 emericella

528	5	31.2	631	10	Q94LM1	Q94LW1 zea mays (m	601	5	31.2	777	5	061366	061366 drosophila
529	5	31.2	635	16	Q8ZLZ2	Q8ZLZ2 salmoneilla	602	5	31.2	789	2	Q93M73	Q93M73 xanthomonas
530	5	31.2	635	16	Q8Z1Y8	Q8Z1Y8 salmoneilla	603	5	31.2	795	5	Q8S565	Q8S565 encephalito
531	5	31.2	637	2	Q9R6L0	Q9R6L0 agrobacteri	604	5	31.2	796	10	Q9L009	Q9L009 arbidopsis
532	5	31.2	637	4	Q8N1Z8	Q8N1Z8 homo sapien	605	5	31.2	797	10	Q6L103	Q6L103 triticum ae
533	5	31.2	637	10	Q94FV9	Q94FV9 abies grand	606	5	31.2	798	5	Q9NEX4	Q9NEX4 caenorhabdi
534	5	31.2	637	16	Q55413	Q55413 gynochocyst	607	5	31.2	801	4	Q9H911	Q9H911 homo sapien
535	5	31.2	639	5	Q21641	Q21641 caenorhabdi	608	5	31.2	801	10	Q9F1G9	Q9F1G9 arbidopsis
536	5	31.2	640	16	Q8YSV0	Q8YSV0 anabaena sp	609	5	31.2	802	16	Q9RUZ9	Q9RUZ9 deinococcus
537	5	31.2	643	5	Q61904	Q61904 caenorhabdi	610	5	31.2	812	5	Q216Z3	Q216Z3 caenorhabdi
538	5	31.2	643	16	Q8YLH4	Q8YLH4 anabaena sp	611	5	31.2	814	13	Q8UTM2	Q8UTM2 oryzae lat
539	5	31.2	643	16	Q8YLH2	Q8YLH2 anabaena sp	612	5	31.2	817	5	Q9VUG1	Q9VUG1 drosophila
540	5	31.2	647	5	Q95R20	Q95R20 babesia ros	613	5	31.2	827	11	Q9QZV2	Q9QZV2 mus musculu
541	5	31.2	651	16	Q9KU86	Q9KU86 vibrio chol	614	5	31.2	827	11	Q91WH4	Q91WH4 mus musculu
542	5	31.2	653	10	Q9FGH0	Q9FGH0 arbidopsis	615	5	31.2	829	10	Q9FLU7	Q9FLU7 arbidopsis
543	5	31.2	653	10	Q9LSA0	Q9LSA0 arbidopsis	616	5	31.2	832	10	Q9PRK1	Q9PRK1 oryza sativ
544	5	31.2	654	3	Q8XIF0	Q8XIF0 cryptococcu	617	5	31.2	832	10	Q91LM3	Q91LM3 oryza sativ
545	5	31.2	657	16	Q99WB3	Q99WB3 staphylococ	618	5	31.2	832	16	Q8XYF4	Q8XYF4 ralestonia s
546	5	31.2	658	3	Q74798	Q74798 schizosacch	619	5	31.2	833	12	Q65480	Q65480 bovine ente
547	5	31.2	659	2	Q9AHF6	Q9AHF6 helicobacte	620	5	31.2	841	11	Q8R2U9	Q8R2U9 mus musculu
548	5	31.2	659	2	Q8RNT6	Q8RNT6 helicobacte	621	5	31.2	842	10	Q91KM2	Q91KM2 oryza sativ
549	5	31.2	659	2	Q8RNT5	Q8RNT5 helicobacte	622	5	31.2	848	2	Q8VVK9	Q8VVK9 corynebacte
550	5	31.2	659	11	Q8R3V2	Q8R3V2 mus musculu	623	5	31.2	859	10	Q9FPS4	Q9FPS4 arbidopsis
551	5	31.2	659	16	Q51552	Q51552 borellia bu	624	5	31.2	863	4	Q9H0P3	Q9H0P3 homo sapien
552	5	31.2	659	16	Q25319	Q25319 helicobacte	625	5	31.2	865	4	Q96K28	Q96K28 homo sapien
553	5	31.2	660	16	Q9ZLN4	Q9ZLN4 helicobacte	626	5	31.2	865	5	Q9N998	Q9N998 leishmania
554	5	31.2	662	2	Q8RLA7	Q8RLA7 pseudomonas	627	5	31.2	870	10	Q8RWQ4	Q8RWQ4 arbidopsis
555	5	31.2	663	10	Q94HT8	Q94HT8 oryza sativ	628	5	31.2	878	5	Q8T6S3	Q8T6S3 giardia lam
556	5	31.2	667	5	Q20932	Q20932 caenorhabdi	629	5	31.2	885	16	Q67173	Q67173 aquilex aeo
557	5	31.2	669	4	Q96M49	Q96M49 homo sapien	630	5	31.2	885	16	Q87658	Q87658 salmoneilla
558	5	31.2	670	2	Q9FCR7	Q9FCR7 neisseria m	631	5	31.2	900	5	Q9V7S2	Q9V7S2 drosophila
559	5	31.2	671	5	Q21046	Q21046 caenorhabdi	632	5	31.2	902	13	Q8UWC5	Q8UWC5 gallus gall
560	5	31.2	673	10	Q9XG31	Q9XG31 guillardia	633	5	31.2	904	5	Q9NBER	Q9NBER leishmania
561	5	31.2	675	10	Q9SRD5	Q9SRD5 aegilops ve	634	5	31.2	908	16	Q8ZRP3	Q8ZRP3 yersteinia pe
562	5	31.2	676	10	Q48785	Q48785 arbidopsis	635	5	31.2	926	10	Q9FVX8	Q9FVX8 arbidopsis
563	5	31.2	681	16	Q8XTU7	Q8XTU7 ralsstonia s	636	5	31.2	927	5	P90947	P90947 caenorhabdi
564	5	31.2	682	2	Q9EXB5	Q9EXB5 neisseria m	637	5	31.2	931	4	Q9HCH8	Q9HCH8 homo sapien
565	5	31.2	683	5	Q9U3G8	Q9U3G8 caenorhabdi	638	5	31.2	939	4	Q9H0B8	Q9H0B8 homo sapien
566	5	31.2	684	2	Q9JPM0	Q9JPM0 neisseria m	639	5	31.2	950	5	Q20733	Q20733 caenorhabdi
567	5	31.2	684	2	Q9FCR6	Q9FCR6 neisseria m	640	5	31.2	963	10	Q8RW88	Q8RW88 arbidopsis
568	5	31.2	686	2	Q9EXB6	Q9EXB6 neisseria m	641	5	31.2	966	11	Q08794	Q08794 mus musculu
569	5	31.2	689	2	Q69750	Q69750 neisseria m	642	5	31.2	975	5	Q19833	Q19833 caenorhabdi
570	5	31.2	690	2	Q33377	Q33377 neisseria g	643	5	31.2	975	16	Q92WC9	Q92WC9 rhizobium m
571	5	31.2	691	16	Q9X6B8	Q9X6B8 caulobacter	644	5	31.2	979	13	Q93509	Q93509 xenopus lae
572	5	31.2	702	3	Q9HPT3	Q9HPT3 neurospora	645	5	31.2	984	5	Q9UOK0	Q9UOK0 plasmodium
573	5	31.2	702	10	Q9CAG9	Q9CAG9 arbidopsis	646	5	31.2	988	4	Q96RMT	Q96RMT homo sapien
574	5	31.2	703	2	Q33376	Q33376 neisseria g	647	5	31.2	990	5	Q9BLV0	Q9BLV0 leishmania
575	5	31.2	711	2	Q9RG26	Q9RG26 staphylococ	648	5	31.2	990	16	Q9KFM7	Q9KFM7 bacillus ha
576	5	31.2	713	3	Q8USV0	Q8USV0 schizosacch	649	5	31.2	997	4	Q9NY16	Q9NY16 homo sapien
577	5	31.2	719	5	Q95R21	Q95R21 babesia ros	650	5	31.2	1007	10	Q9ZVD4	Q9ZVD4 arbidopsis
578	5	31.2	720	8	Q9RTA1	Q9RTA1 matelea bif	651	5	31.2	1009	5	Q9VAT3	Q9VAT3 drosophila
579	5	31.2	723	11	Q9D677	Q9D677 mus musculu	652	5	31.2	1015	10	Q22243	Q22243 arbidopsis
580	5	31.2	724	16	Q9J8C9	Q9J8C9 spodoptera	653	5	31.2	1016	15	Q91NV0	Q91NV0 human immun
581	5	31.2	727	12	Q92XG8	Q92XG8 rhizobium m	654	5	31.2	1023	10	Q8RYG4	Q8RYG4 oryza sativ
582	5	31.2	727	16	Q8YP24	Q8YP24 anabaena sp	655	5	31.2	1025	15	Q91WZ6	Q91WZ6 human immun
583	5	31.2	731	16	Q8RG00	Q8RG00 fusobacteri	656	5	31.2	1027	5	Q8SX14	Q8SX14 drosophila
584	5	31.2	732	2	Q8VP74	Q8VP74 pseudomonas	657	5	31.2	1031	4	Q9BY57	Q9BY57 homo sapien
585	5	31.2	739	13	Q9FTJ3	Q9FTJ3 oncorhynch	658	5	31.2	1034	16	Q931B0	Q931B0 rhizobium m
586	5	31.2	739	13	P79908	P79908 salmo salar	659	5	31.2	1045	10	Q8R201	Q8R201 oryza sativ
587	5	31.2	742	11	Q9TLA2	Q9TLA2 araujia ser	660	5	31.2	1051	5	Q9VSE2	Q9VSE2 drosophila
588	5	31.2	742	11	Q9CWS4	Q9CWS4 mus musculu	661	5	31.2	1055	16	Q98B14	Q98B14 rhizobium 1
589	5	31.2	744	16	Q33087	Q33087 mycobacteri	662	5	31.2	1056	11	Q9Z0W6	Q9Z0W6 mus musculu
590	5	31.2	750	16	Q8UG82	Q8UG82 agrobacteri	663	5	31.2	1064	11	Q91Y58	Q91Y58 mus musculu
591	5	31.2	755	11	Q9R0G6	Q9R0G6 mus musculu	664	5	31.2	1070	5	Q97203	Q97203 leishmania
592	5	31.2	755	11	Q8V154	Q8V154 mus musculu	665	5	31.2	1077	10	Q9FX91	Q9FX91 arbidopsis
593	5	31.2	764	16	Q92A22	Q92A22 listeria in	666	5	31.2	1080	10	Q8VY06	Q8VY06 arbidopsis
594	5	31.2	765	5	Q9GPL5	Q9GPL5 drosophila	667	5	31.2	1091	11	Q9Z1L5	Q9Z1L5 mus musculu
595	5	31.2	765	5	Q9G6K1	Q9G6K1 drosophila	668	5	31.2	1095	10	Q94J79	Q94J79 oryza sativ
596	5	31.2	765	16	Q91054	Q91054 streptomyce	669	5	31.2	1099	10	Q943V7	Q943V7 oryza sativ
597	5	31.2	773	2	Q9XBM4	Q9XBM4 myxococcus	670	5	31.2	1099	16	Q8HYC2	Q8HYC2 pseudomonas
598	5	31.2	774	16	Q9X6W0	Q9X6W0 pseudomonas	671	5	31.2	1105	4	Q8TBV4	Q8TBV4 homo sapien
599	5	31.2	775	12	Q56290	Q56290 human herpe	672	5	31.2	1105	10	Q93VN8	Q93VN8 phaeoculus v
600	5	31.2	775	16	Q9CNH5	Q9CNH5 pasteurella	673	5	31.2	1110	10	Q9F177	Q9F177 arbidopsis

674	5	31.2	1114	5	Q27811	tritpneustes	747	5	31.2	2023	5	Q9V529	Q9V529 drosophila
675	5	31.2	1118	10	Q93VL6	Q93VL6 phaseolus v	748	5	31.2	2155	4	Q75443	Q75443 homo sapien
676	5	31.2	1120	5	Q08773	Q08773 saeccharomyc	749	5	31.2	2155	11	Q08523	Q08523 mus musculus
677	5	31.2	1126	5	Q9VGK5	Q9VGK5 drosophila	750	5	31.2	2194	12	Q914E1	Q914E1 human echov
678	5	31.2	1133	10	Q93VS9	Q93VS9 phaseolus v	751	5	31.2	2194	12	Q914E0	Q914E0 human echov
679	5	31.2	1154	11	Q9QVN3	Q9QVN3 ratius sp.	752	5	31.2	2216	10	Q9LV03	Q9LV03 arabisdopsis
680	5	31.2	1165	13	Q9YH40	Q9YH40 xiphophorus	753	5	31.2	2258	12	Q89238	Q89238 human parat
681	5	31.2	1166	11	Q9QVN4	Q9QVN4 ratius sp.	754	5	31.2	2258	12	Q81081	Q81081 human parat
682	5	31.2	1168	10	Q9LE81	Q9LE81 arabisdopsis	755	5	31.2	2266	4	Q8WYPS	Q8WYPS homo sapien
683	5	31.2	1169	5	Q96055	Q96055 anthocidari	756	5	31.2	2369	3	Q96VL6	Q96VL6 candida alb
684	5	31.2	1180	4	Q15051	Q15051 homo sapien	757	5	31.2	2392	4	Q95001	Q95001 homo sapien
685	5	31.2	1181	5	Q9U308	Q9U308 caenorhabdi	758	5	31.2	2412	4	Q92616	Q92616 homo sapien
686	5	31.2	1181	10	Q9LV05	Q9LV05 arabisdopsis	759	5	31.2	2485	5	Q96134	Q96134 plasmodium
687	5	31.2	1190	4	Q9FNB4	Q9FNB4 arabisdopsis	760	5	31.2	2516	5	Q9V3P7	Q9V3P7 drosophila
688	5	31.2	1204	4	Q96Q83	Q96Q83 homo sapien	761	5	31.2	2566	16	Q8U9P3	Q8U9P3 agrobacteri
689	5	31.2	1204	5	Q18605	Q18605 caenorhabdi	762	5	31.2	2829	13	P70039	P70039 xenopus lae
690	5	31.2	1205	11	P97686	P97686 ratius norv	763	5	31.2	2870	5	Q77380	Q77380 plasmodium
691	5	31.2	1217	10	Q9SCX7	Q9SCX7 arabisdopsis	764	5	31.2	3105	12	Q8U219	Q8U219 cercopithec
692	5	31.2	1217	10	Q9SCX6	Q9SCX6 arabisdopsis	765	5	31.2	3229	5	Q26912	Q26912 trypanosoma
693	5	31.2	1217	10	Q9XGM3	Q9XGM3 arabisdopsis	766	5	31.2	3487	2	Q9FDU0	Q9FDU0 microcytis
694	5	31.2	1236	4	Q9UHI3	Q9UHI3 homo sapien	767	5	31.2	3487	2	Q9RNB3	Q9RNB3 microcytis
695	5	31.2	1244	4	Q8TEW1	Q8TEW1 homo sapien	768	5	31.2	3524	2	Q93HB6	Q93HB6 streptomyc
696	5	31.2	1245	4	Q9ULD2	Q9ULD2 homo sapien	769	5	31.2	4498	13	Q93291	Q93291 fuqu rubrip
697	5	31.2	1250	5	Q27813	Q27813 tritpneustes	770	5	31.2	4835	5	Q8T5T1	Q8T5T1 giardia lam
698	5	31.2	1263	10	Q48686	Q48686 arabisdopsis	771	5	31.2	6307	4	Q8WKG9	Q8WKG9 homo sapien
699	5	31.2	1266	4	Q9NYE6	Q9NYE6 homo sapien	772	5	31.2	6831	5	Q23550	Q23550 caenorhabdi
700	5	31.2	1273	4	Q96RM6	Q96RM6 homo sapien	773	5	31.2	7160	5	Q23551	Q23551 caenorhabdi
701	5	31.2	1274	10	Q9M680	Q9M680 dunaliella	774	5	31.2	8243	8	Q96554	Q96554 cryptospori
702	5	31.2	1274	10	P93125	P93125 dunaliella	775	5	31.2	8	11	Q9OVJ8	Q9OVJ8 mus sp. mep
703	5	31.2	1275	12	Q8QVF3	Q8QVF3 eyach vltus	776	5	31.2	16	4	Q9UCR2	Q9UCR2 homo sapien
704	5	31.2	1292	5	Q95N19	Q95N19 plasmodium	777	5	31.2	18	2	P72015	P72015 mycobacteri
705	5	31.2	1299	4	Q15179	Q15179 homo sapien	778	5	31.2	21	2	Q9RSU9	Q9RSU9 rhodococcus
706	5	31.2	1299	4	Q92823	Q92823 homo sapien	779	5	31.2	22	3	P82261	P82261 aspergillus
707	5	31.2	1308	4	Q9UHI4	Q9UHI4 homo sapien	780	5	31.2	24	2	Q9R3J2	Q9R3J2 ureaplasma
708	5	31.2	1310	4	Q8TEW0	Q8TEW0 homo sapien	781	5	31.2	24	8	Q95CT3	Q95CT3 atlingia ex
709	5	31.2	1313	3	Q8X0S9	Q8X0S9 neurospora	782	5	31.2	24	8	Q95CT1	Q95CT1 atlingia ta
710	5	31.2	1319	4	Q8TEW2	Q8TEW2 homo sapien	783	5	31.2	24	8	Q95CH9	Q95CH9 atlingia gr
711	5	31.2	1333	11	Q99NH2	Q99NH2 mus musculu	784	5	31.2	24	8	Q95CH7	Q95CH7 semilliquida
712	5	31.2	1337	11	Q92340	Q92340 ratius norv	785	5	31.2	24	8	Q95CH5	Q95CH5 semilliquida
713	5	31.2	1353	4	Q9BY56	Q9BY56 homo sapien	786	5	31.2	24	8	Q95CH3	Q95CH3 liquidambar
714	5	31.2	1356	4	Q8TEW3	Q8TEW3 homo sapien	787	5	31.2	24	8	Q95CH1	Q95CH1 liquidambar
715	5	31.2	1360	13	Q9YGS5	Q9YGS5 xenopus lae	788	5	31.2	24	8	Q95CG9	Q95CG9 atlingia ch
716	5	31.2	1364	13	Q9YHY6	Q9YHY6 xenopus lae	789	5	31.2	24	8	Q95CG7	Q95CG7 hamamelis j
717	5	31.2	1411	2	Q9AJ93	Q9AJ93 actinomycetes	790	5	31.2	24	8	Q95CG5	Q95CG5 mytilaria l
718	5	31.2	1414	3	Q96WQ2	Q96WQ2 penicillium	791	5	31.2	24	8	Q94PI1	Q94PI1 liquidambar
719	5	31.2	1438	3	Q06681	Q06681 saeccharomyc	792	5	31.2	24	8	Q94PI0	Q94PI0 liquidambar
720	5	31.2	1449	5	Q10922	Q10922 caenorhabdi	793	5	31.2	24	8	Q94P09	Q94P09 atlingia ob
721	5	31.2	1467	16	Q62334	Q62334 caenorhabdi	794	5	31.2	25	6	Q95UJ3	Q95UJ3 sus scrofa
722	5	31.2	1478	16	Q9PDJ6	Q9PDJ6 xylella fas	795	5	31.2	26	15	Q9JFP1	Q9JFP1 human t-lym
723	5	31.2	1479	10	Q9ATV5	Q9ATV5 arabisdopsis	796	5	31.2	27	12	Q89139	Q89139 hepatitis c
724	5	31.2	1518	17	Q8THT8	Q8THT8 methanosarc	797	5	31.2	28	2	Q93MG3	Q93MG3 thiodactiliu
725	5	31.2	1522	10	Q9LRT2	Q9LRT2 arabisdopsis	798	5	31.2	29	6	Q9N1V3	Q9N1V3 equus caball
726	5	31.2	1552	2	Q9ATU6	Q9ATU6 neisseria m	799	5	31.2	29	8	Q9T1G1	Q9T1G1 allonyncarp
727	5	31.2	1561	2	Q51169	Q51169 neisseria m	800	5	31.2	29	16	Q86497	Q86497 streptomyc
728	5	31.2	1581	16	Q8U9E4	Q8U9E4 agrobacteri	801	5	31.2	30	10	P81716	P81716 schistosom
729	5	31.2	1635	5	Q17368	Q17368 caenorhabdi	802	5	31.2	30	10	Q9S9G5	Q9S9G5 hellianthus
730	5	31.2	1637	10	Q8S1N7	Q8S1N7 oryza sativ	803	5	31.2	31	2	Q9R4X1	Q9R4X1 treponema d
731	5	31.2	1650	5	Q45176	Q45176 caenorhabdi	804	5	31.2	31	4	Q96058	Q96058 homo sapien
732	5	31.2	1672	16	Q8Y3C6	Q8Y3C6 ralsstonia s	805	5	31.2	33	5	Q25159	Q25159 haliclona s
733	5	31.2	1674	10	Q64923	Q64923 zea mays (m	806	5	31.2	33	13	Q8U0R8	Q8U0R8 brachydanio
734	5	31.2	1706	3	Q42900	Q42900 schizosacch	807	5	31.2	33	16	Q9A6D5	Q9A6D5 caulobacter
735	5	31.2	1771	16	P74647	P74647 synecocyst	808	5	31.2	34	6	Q9TQZ4	Q9TQZ4 sus scrofa
736	5	31.2	1773	16	Q9JVB9	Q9JVB9 neisseria m	809	5	31.2	34	11	Q91XZ3	Q91XZ3 ratius norv
737	5	31.2	1785	5	Q25685	Q25685 plasmodium	810	5	31.2	35	2	Q9Z9H9	Q9Z9H9 magnetospir
738	5	31.2	1807	4	Q9ULH6	Q9ULH6 homo sapien	811	5	31.2	36	5	Q26277	Q26277 drosophila
739	5	31.2	1815	16	Q9K0B4	Q9K0B4 neisseria m	812	5	31.2	36	16	Q93RS8	Q93RS8 streptomyc
740	5	31.2	1819	10	Q9SGP0	Q9SGP0 arabisdopsis	813	5	31.2	37	12	Q9IM08	Q9IM08 bovine help
741	5	31.2	1858	3	Q9Y7B3	Q9Y7B3 emeritella	814	5	31.2	38	11	Q9E8S9	Q9E8S9 ratius norv
742	5	31.2	1882	3	Q9P898	Q9P898 podospora a	815	5	31.2	38	11	Q92500	Q92500 cisticulus
743	5	31.2	1887	3	Q9HGP6	Q9HGP6 schizosacch	816	5	31.2	39	13	Q9P803	Q9P803 naja naja (
744	5	31.2	1928	4	Q99736	Q99736 homo sapien	817	5	31.2	40	11	Q8R2N7	Q8R2N7 mus musculu
745	5	31.2	2016	5	Q9VG78	Q9VG78 drosophila	818	5	31.2	40	12	Q86958	Q86958 human herpe
746	5	31.2	2023	5	Q96542	Q96542 drosophila	819	5	31.2	40	12	Q91X00	Q91X00 human coxa



820	4	25.0	40	12	Q91W29	Q91w29 human echov	893	4	25.0	59	16	Q98UJ1	Q98jui rhizobium 1
821	4	25.0	40	12	Q91W28	Q91w28 human coxsaa	894	4	25.0	60	12	Q9EMV5	Q9emv5 amasacta moo
822	4	25.0	40	12	Q91W27	Q91w27 human coxsaa	895	4	25.0	61	10	Q94ZU8	Q94zu8 oryza sativ
823	4	25.0	40	12	Q91W26	Q91w26 human coxsaa	896	4	25.0	61	13	Q9DFR0	Q9dfc0 gilllichthys
824	4	25.0	40	12	Q91W25	Q91w25 human echov	897	4	25.0	61	16	Q9A4Y2	Q9a4y2 caulobacter
825	4	25.0	40	12	Q91W21	Q91w21 human coxsaa	898	4	25.0	62	5	Q45226	Q45226 caenorhabdi
826	4	25.0	40	12	Q91W20	Q91w20 human coxsaa	899	4	25.0	62	6	Q9M2Z1	Q9m2zi macaca fasc
827	4	25.0	40	12	Q91W19	Q91w19 human echov	900	4	25.0	62	6	Q28581	Q28581 ovis aries
828	4	25.0	40	12	Q91W13	Q91w13 human echov	901	4	25.0	62	10	Q41956	Q41956 arabisdopsis
829	4	25.0	40	12	Q91W12	Q91w12 human echov	902	4	25.0	62	10	Q92745	Q92745 arabisdopsis
830	4	25.0	40	12	Q91W11	Q91w11 human echov	903	4	25.0	62	11	Q08632	Q08632 mus musculu
831	4	25.0	40	12	Q91W10	Q91w10 human echov	904	4	25.0	62	15	Q11107	Q11107 human immun
832	4	25.0	40	12	Q91W09	Q91w09 human echov	905	4	25.0	62	15	Q11110	Q11110 human immun
833	4	25.0	40	12	Q91WX8	Q91wx8 human echov	906	4	25.0	62	16	Q9PEU8	Q9peu8 xylella fas
834	4	25.0	40	12	Q91WX6	Q91wx6 human coxsaa	907	4	25.0	62	16	Q98BP6	Q98bp6 rhizobium 1
835	4	25.0	40	12	Q91WX5	Q91wx5 human enter	908	4	25.0	62	17	Q9UYD5	Q9uyd5 pyrococcus
836	4	25.0	41	16	Q9UDP4	Q9uda4 homo sapien	909	4	25.0	62	17	Q9HGV0	Q9hgv0 halobacteri
837	4	25.0	41	16	Q9WXL6	Q9wxi6 thermotoga	910	4	25.0	63	5	Q95PA2	Q95pa2 bradyelia hy
838	4	25.0	42	5	Q61336	Q61336 panulirus 1	911	4	25.0	63	10	Q9SA97	Q9sa97 arabisdopsis
839	4	25.0	43	11	Q91ZA2	Q91za2 meriones un	912	4	25.0	63	10	Q93WS5	Q93ws5 triticum ae
840	4	25.0	44	6	Q02774	Q02774 sus scrofa	913	4	25.0	63	12	Q88988	Q88988 variola vir
841	4	25.0	44	15	Q9OCX1	Q9ocx1 human t-lym	914	4	25.0	63	12	Q89477	Q89477 variola vir
842	4	25.0	44	15	Q10490	Q10490 human immun	915	4	25.0	64	2	Q9XD17	Q9xd17 streptococc
843	4	25.0	47	13	Q90774	Q90774 gallus gall	916	4	25.0	64	4	Q9UD19	Q9udi9 homo sapien
844	4	25.0	48	6	Q28750	Q28750 ovis sp. hg	917	4	25.0	64	10	Q9FNT4	Q9fnt4 braesica ol
845	4	25.0	49	17	Q9HHB5	Q9hhe5 halobacteri	918	4	25.0	64	11	Q62675	Q62675 rattus norv
846	4	25.0	50	12	Q66827	Q66827 human echov	919	4	25.0	64	13	Q91220	Q91220 oncorhynch
847	4	25.0	50	12	Q66828	Q66828 human echov	920	4	25.0	64	15	Q85624	Q85624 avian myelo
848	4	25.0	50	12	Q66789	Q66789 human echov	921	4	25.0	64	15	Q85482	Q85482 rous sarcom
849	4	25.0	50	12	Q66792	Q66792 human echov	922	4	25.0	64	15	Q85470	Q85470 rous sarcom
850	4	25.0	50	12	Q66796	Q66796 human echov	923	4	25.0	64	16	Q9KSA8	Q9ksa8 vibrio chol
851	4	25.0	50	12	Q66797	Q66797 human echov	924	4	25.0	64	16	Q9CDV6	Q9cdv6 lactococcus
852	4	25.0	50	12	Q66815	Q66815 human echov	925	4	25.0	64	16	Q9C116	Q9c116 lactococcus
853	4	25.0	50	12	Q66822	Q66822 human echov	926	4	25.0	65	1	Q93702	Q93702 sulfolobus
854	4	25.0	50	12	Q66823	Q66823 human echov	927	4	25.0	65	5	Q44665	Q44665 caenorhabdi
855	4	25.0	50	12	Q9WLG3	Q9wlg3 human coxsaa	928	4	25.0	65	11	P70631	P70631 rattus norv
856	4	25.0	50	12	Q9WLG2	Q9wlg2 human echov	929	4	25.0	65	16	Q984E4	Q984e4 rhizobium 1
857	4	25.0	50	12	Q9YID7	Q9yid7 human echov	930	4	25.0	65	17	Q9HNL9	Q9hnl9 halobacteri
858	4	25.0	50	12	Q9YUV5	Q9yuv5 human echov	931	4	25.0	65	17	Q73974	Q73974 pyrococcus
859	4	25.0	50	12	Q9YUV3	Q9yuv3 human coxsaa	932	4	25.0	66	2	Q9X6K9	Q9x6k9 klebsiella
860	4	25.0	50	12	Q9UN57	Q9un57 human echov	933	4	25.0	66	2	Q32951	Q32951 mycobacteri
861	4	25.0	50	12	Q66817	Q66817 human echov	934	4	25.0	66	10	Q93WS8	Q93ws8 triticum ae
862	4	25.0	51	8	Q92ZV8	Q92zv8 saccharomyc	935	4	25.0	66	12	Q91DB4	Q91db4 hepatitis c
863	4	25.0	51	12	Q66825	Q66825 human echov	936	4	25.0	66	13	P87483	P87483 gallus gall
864	4	25.0	51	12	Q66848	Q66848 human echov	937	4	25.0	66	13	Q90217	Q90217 abraxas bra
865	4	25.0	51	17	Q97YV3	Q97yv3 sulfolobus	938	4	25.0	66	16	Q8RYN0	Q8ryn0 anabaena sp
866	4	25.0	52	12	Q09728	Q09728 leonurus mo	939	4	25.0	66	16	Q8R308	Q8r308 thermoaer
867	4	25.0	52	17	Q971S4	Q971s4 sulfolobus	940	4	25.0	67	2	Q9F1H8	Q9f1h8 enterococcu
868	4	25.0	53	2	Q9F265	Q9f265 actinobacil	941	4	25.0	67	3	Q04334	Q04334 pichia angu
869	4	25.0	53	4	Q9P138	Q9p138 homo sapien	942	4	25.0	67	4	Q9BUD3	Q9bud3 homo sapien
870	4	25.0	53	11	Q08633	Q08633 mus musculu	943	4	25.0	67	4	Q9SMG1	Q9smg1 secale cere
871	4	25.0	53	16	Q8XNY2	Q8xny2 clostridium	944	4	25.0	67	10	Q9SEB7	Q9seb7 secale scir
872	4	25.0	53	17	Q979Q9	Q979q9 thermoplas	945	4	25.0	67	10	Q93WS7	Q93ws7 triticum ae
873	4	25.0	54	6	Q95M43	Q95m43 bos taurus	946	4	25.0	67	10	Q93WS7	Q93ws7 triticum ae
874	4	25.0	54	10	Q41961	Q41961 arabisdopsis	947	4	25.0	67	10	Q93WS6	Q93ws6 paramedum
875	4	25.0	55	4	Q9UC78	Q9uc78 homo sapien	948	4	25.0	67	12	Q98560	Q98560 paramedum
876	4	25.0	55	5	P83282	P83282 cichada flam	949	4	25.0	67	12	Q981J1	Q981j1 sin nombr
877	4	25.0	55	12	Q91LA6	Q91la6 white spot	950	4	25.0	67	13	Q92165	Q92165 oncorhynch
878	4	25.0	55	12	Q8QPU5	Q8qpu5 tomato seve	951	4	25.0	67	13	Q91235	Q91235 oncorhynch
879	4	25.0	56	5	Q9VB35	Q9vb35 diosophila	952	4	25.0	67	13	Q91172	Q91172 oncorhynch
880	4	25.0	56	10	Q42242	Q42242 arabisdopsis	953	4	25.0	67	13	Q91158	Q91158 oncorhynch
881	4	25.0	57	2	P71460	P71460 lactobacill	954	4	25.0	67	13	Q92166	Q92166 oncorhynch
882	4	25.0	57	5	Q9NMG4	Q9nmg4 leishmania	955	4	25.0	67	16	Q53937	Q53937 mycobacteri
883	4	25.0	57	12	Q9YR30	Q9yr30 tt virus. h	956	4	25.0	68	1	Q50780	Q50780 methanobact
884	4	25.0	57	16	Q50838	Q50838 borellia bu	957	4	25.0	68	3	Q9URB4	Q9urb4 saccharomyc
885	4	25.0	58	2	Q9X4E7	Q9x4e7 rhodobacter	958	4	25.0	68	10	Q98992	Q98992 braesica na
886	4	25.0	58	5	Q9NTC1	Q9ntc1 leishmania	959	4	25.0	68	12	Q67546	Q67546 camelpox vi
887	4	25.0	58	9	Q9T204	Q9t204 lactococcus	960	4	25.0	68	12	Q8V216	Q8v216 camelpox vi
888	4	25.0	58	9	Q8W6W1	Q8w6w1 cyanophage	961	4	25.0	68	16	Q981U6	Q981u6 rhizobium 1
889	4	25.0	58	12	Q8V6T4	Q8v6t4 halovirus h	962	4	25.0	69	5	Q44663	Q44663 caenorhabdi
890	4	25.0	58	16	Q92HA4	Q92ha4 ticketelia	963	4	25.0	69	5	Q96431	Q96431 babesia cab
891	4	25.0	58	17	Q8ZWE7	Q8zwe7 pyrobaculum	964	4	25.0	69	6	Q46476	Q46476 felis silve
892	4	25.0	59	10	Q9S913	Q9s913 arabisdopsis	965	4	25.0	69	9	Q38001	Q38001 bacterioph

966	4	25.0	69	12	Q68373	Q68373 hepatitis c
967	4	25.0	69	12	Q8QWD8	Q8QWD8 human echov
968	4	25.0	69	16	Q9EWV8	Q9EWV8 streptomyces
969	4	25.0	69	17	Q8THS3	Q8THS3 methanocarc
970	4	25.0	70	5	Q9N7S6	Q9N7S6 leishmania
971	4	25.0	70	10	Q9ZMU0	Q9ZMU0 glycine max
972	4	25.0	70	10	Q04227	Q04227 plantago ma
973	4	25.0	70	10	Q8S2Q2	Q8S2Q2 oryza sativ
974	4	25.0	70	11	Q9QVU9	Q9QVU9 mus sp. hox
975	4	25.0	70	12	Q66477	Q66477 human enter
976	4	25.0	70	12	Q9WKC6	Q9WKC6 hepatitis b
977	4	25.0	70	13	Q9PRL1	Q9PRL1 gallus sp.
978	4	25.0	70	16	Q9K9J0	Q9K9J0 bacillus ha
979	4	25.0	70	16	Q98UB6	Q98UB6 rhizobium 1
980	4	25.0	70	16	Q8UCX6	Q8UCX6 agrobacteri
981	4	25.0	71	4	Q13313	Q13313 homo sapien
982	4	25.0	71	12	Q84429	Q84429 paramedium
983	4	25.0	71	12	Q9EMP5	Q9EMP5 amasacta moo
984	4	25.0	71	16	P73164	P73164 synecocyst
985	4	25.0	71	16	Q8YFD5	Q8YFD5 brucella me
986	4	25.0	71	16	Q8UD71	Q8UD71 agrobacteri
987	4	25.0	71	17	Q9YA12	Q9YA12 aeropyrum p
988	4	25.0	72	5	Q8XWV3	Q8XWV3 caenorhabdi
989	4	25.0	72	5	Q27398	Q27398 loligo peal
990	4	25.0	72	5	Q8T951	Q8T951 drosophila
991	4	25.0	72	9	Q94MT5	Q94MT5 bacterioph
992	4	25.0	72	10	Q04241	Q04241 zea mays (m
993	4	25.0	72	10	Q04236	Q04236 vicia faba
994	4	25.0	72	12	Q91RV2	Q91RV2 lettuce mos
995	4	25.0	72	12	Q91RV0	Q91RV0 lettuce mos
996	4	25.0	72	12	Q91R08	Q91R08 lettuce mos
997	4	25.0	72	12	Q912F1	Q912F1 human echov
998	4	25.0	72	12	Q8VB73	Q8VB73 white spot
999	4	25.0	72	16	Q92JF0	Q92JF0 rickettsia
1000	4	25.0	72	16	Q8UCE1	Q8UCE1 agrobacteri

## ALIGNMENTS

## RESULT 1

Q26552 ID Q26552 PRELIMINARY; PRT; 274 AA.  
 AC Q26552;  
 DT 01-NOV-1996 (TREMBLrel. 01, Created)  
 DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)  
 DT 01-MAR-2002 (TREMBLrel. 20, Last annotation update)  
 DE Elastase.  
 OS Schistosoma mansoni (Blood fluke).  
 OC Eukaryota; Metazoa; Platyhelminthes; Trematoda; Digenea; Strigeidida;  
 OC Schistosomatidae; Schistosomatidae; Schistosoma.  
 OX NCBI\_TaxID=6183;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=PUERTO RICO;  
 RX MEDLINE=96362066; PubMed=8720180;  
 RA Pierrot C., Capron A., Khalife J.;  
 RT "Cloning and characterization of two genes encoding Schistosoma  
 mansoni elastase.";  
 RL Mol. Biochem. Parasitol. 75:113-117(1995).  
 CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE  
 TRYPSIN FAMILY.  
 DR EMBL: U31768; AAC46967.1; -.  
 DR HSSP: P20231; IAMO.  
 DR MEROPS: S01.144; -.  
 DR InterPro: IPR001314; Chymotrypsin.  
 DR InterPro: IPR001254; Ser\_protease\_Try.  
 DR Pfam: PF00089; trypsin.1.  
 DR PRINTS: PR00722; CHYMOTRYPSIN.  
 DR SMART: SM00020; Tryp\_Spc.1.  
 DR PROSITE: PS0240; TRYPSIN\_DOM.1.  
 DR PROSITE: PS00134; TRYPSIN\_HIS; UNKNOWN\_1.  
 KW Hydrolyase; Serine protease.

SEQ SEQUENCE 274 AA; 29586 MW; 53F21914FDD3DA450 CRC64;  
 Query Match 100.0%; Score 16; DB 5; Length 274;  
 Best Local Similarity 100.0%; Pred. No. 3,4e-10;  
 Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 VGYGRDNDPSPSRN 16  
 |||||  
 Db 168 VGYGRDNDPSPSRN 183

## RESULT 2

Q26553 ID Q26553 PRELIMINARY; PRT; 257 AA.

AC Q26553;  
 DT 01-NOV-1996 (TREMBLrel. 01, Created)  
 DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)  
 DT 01-MAR-2002 (TREMBLrel. 20, Last annotation update)  
 DE Elastase.  
 OS Schistosoma mansoni (Blood fluke).  
 OC Eukaryota; Metazoa; Platyhelminthes; Trematoda; Digenea; Strigeidida;  
 OC Schistosomatidae; Schistosomatidae; Schistosoma.  
 OX NCBI\_TaxID=6183;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=PUERTO RICO;  
 RX MEDLINE=96362066; PubMed=8720180;  
 RA Pierrot C., Capron A., Khalife J.;  
 RT "Cloning and characterization of two genes encoding Schistosoma  
 mansoni elastase.";  
 RL Mol. Biochem. Parasitol. 75:113-117(1995).  
 DR EMBL: U31769; AAC46968.1; -.  
 DR HSSP: P00766; 1GCD.  
 DR InterPro: IPR001254; Ser\_protease\_Try.  
 DR Pfam: PF00089; trypsin.1.  
 DR SMART: SM00020; Tryp\_Spc.1.  
 DR PROSITE: PS0240; TRYPSIN\_DOM.1.  
 DR PROSITE: PS00134; TRYPSIN\_HIS; UNKNOWN\_1.  
 KW Hydrolyase; Serine protease.

SEQ SEQUENCE 257 AA; 27971 MW; 00F8872799F9DD64 CRC64;

Query Match 87.5%; Score 14; DB 5; Length 257;  
 Best Local Similarity 100.0%; Pred. No. 5,4e-08;  
 Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 VGYGRDNDPSPSR 14  
 |||||  
 Db 164 VGYGRDNDPSPSR 177

## RESULT 3

Q9A6Y0 ID Q9A6Y0 PRELIMINARY; PRT; 686 AA.

AC Q9A6Y0;  
 DT 01-JUN-2001 (TREMBLrel. 17, Created)  
 DT 01-JUN-2001 (TREMBLrel. 17, Last sequence update)  
 DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)  
 DE NADH dehydrogenase I, G subunit.  
 GN CC1946.  
 OS Caulobacter crescentus.  
 OC Bacteria; Proteobacteria; alpha subdivision; Caulobacter group;  
 OC Caulobacter.  
 OX NCBI\_TaxID=155892;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=ATCC 19089 / CB15;  
 RX MEDLINE=21173698; PubMed=11259647;  
 RA Eiten J., Heidelberg J.F., Alley M.R.K., Ohta N., Maddock J.R.,  
 RA Pirocka I., Nelson W.C., Newton A., Stephens C., Phadke N.D., Ely B.,  
 RA Deboy R.T., Dodson R.J., Durkin A.S., Gaim M.L., Hatt D.H.,  
 RA Kolonay J.F., Smit J., Craven M.B., Khouri H., Shetty J., Berry K.,  
 RA Uterback T., Tran K., Wolf A., Vamathevan J., Ermolaeva M., White O.,

RA Salzberg S.L., Venter J.C., Shapiro L., Fraser C.M.;  
 RT "Complete genome sequence of *Caulobacter crescentus*,"  
 RL Proc. Natl. Acad. Sci. U.S.A. 98:4136-4141(2001).  
 DR EMBL; AE005868; AAK23921.1; -.  
 DR TIGR; CC1946; -.  
 DR InterPro; IPR000283; Complex1\_75K.  
 DR InterPro; IPR001041; Perletoxlin.  
 DR InterPro; IPR001467; Prot\_Mboxred.  
 DR Pfam; PF00111; fer2; 1.  
 DR Pfam; PF00384; molybdopterin; 1.  
 DR PROSITE; PS00641; COMPLEX1\_75K\_1; 1.  
 DR PROSITE; PS00642; COMPLEX1\_75K\_2; 1.  
 DR PROSITE; PS00643; COMPLEX1\_75K\_3; 1.  
 KM Iron-sulfur; Complete proteome.  
 SQ SEQUENCE 686 AA; 73163 MW; 3B4955F60FB55F70 CRC64;

Query Match 43.8%; Score 7; DB 16; Length 686;  
 Best Local Similarity 100.0%; Pred. No. 7.5;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 VGYGRD 7  
 |||||  
 DB 115 VGYGRD 121

RESULT 4  
 O919J2 PRELIMINARY; PRT; 142 AA.  
 AC O919J2;  
 DT 01-OCT-2000 (TRENBLrel. 15, Created)  
 DT 01-OCT-2000 (TRENBLrel. 15, Last sequence update)  
 DT 01-DEC-2001 (TRENBLrel. 19, Last annotation update)  
 DE P75 neurotrophin receptor b (Fragment).  
 GN P75NTRB  
 OS Xenopus laevis (African clawed frog).  
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
 OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipidea; Pipidae;  
 OC Xenopodinae; Xenopus.  
 CX NCBI\_TaxID=8355;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Hutson L.D., Bothwell M.;  
 RT "Developmental expression of two putative p75NTR genes in *Xenopus laevis*,"  
 RT Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.  
 RL EMBL; AF246463; AAF63523.1; -.  
 DR HSSP; O14763; IDOG.  
 DR InterPro; IPR001368; TNFR\_c6.  
 DR Pfam; PF00020; TNFR\_c6; 2.  
 DR SMART; SM00208; TNFR; 2.  
 DR PROSITE; PS00652; TNFR\_NGFR\_1; UNKNOWN\_1.  
 DR PROSITE; PS50050; TNFR\_NGFR\_2; 1.  
 KW Receptor.  
 FT NON\_TER  
 FT NON\_TER  
 SQ SEQUENCE 142 AA; 15120 MW; 3C50CFE7DD1B57FD CRC64;

Query Match 37.5%; Score 6; DB 13; Length 142;  
 Best Local Similarity 100.0%; Pred. No. 24;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 7 DNDRDP 12  
 |||||  
 DB 44 DNDRDP 49

RESULT 5  
 O919J3 PRELIMINARY; PRT; 143 AA.  
 AC O919J3;  
 DT 01-OCT-2000 (TRENBLrel. 15, Created)  
 DT 01-OCT-2000 (TRENBLrel. 15, Last sequence update)  
 DT 01-DEC-2001 (TRENBLrel. 19, Last annotation update)

OY 2 GYGRD 7  
 |||||  
 DB 102 GYGRD 107

DE P75 neurotrophin receptor a (Fragment).  
 GN P75NTRA.  
 OS Xenopus laevis (African clawed frog).  
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
 OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipidea; Pipidae;  
 OC Xenopodinae; Xenopus.  
 CX NCBI\_TaxID=8355;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Hutson L.D., Bothwell M.;  
 RT "Developmental expression of two putative p75NTR genes in *Xenopus laevis*,"  
 RT Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.  
 RL EMBL; AF246462; AAF63522.1; -.  
 DR HSSP; P19438; IEXT.  
 DR InterPro; IPR001899; Gram\_pos\_anchor.  
 DR InterPro; IPR001368; TNFR\_c6.  
 DR Pfam; PF00020; TNFR\_c6; 2.  
 DR SMART; SM00208; TNFR; 2.  
 DR PROSITE; PS00343; GRAM\_POS\_ANCHORING; UNKNOWN\_1.  
 DR PROSITE; PS00652; TNFR\_NGFR\_1; UNKNOWN\_1.  
 DR PROSITE; PS50050; TNFR\_NGFR\_2; 1.  
 KW Receptor.  
 FT NON\_TER  
 FT NON\_TER  
 SQ SEQUENCE 143 AA; 15031 MW; D583924B847AC324 CRC64;

Query Match 37.5%; Score 6; DB 13; Length 143;  
 Best Local Similarity 100.0%; Pred. No. 24;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 7 DNDRDP 12  
 |||||  
 DB 44 DNDRDP 49

RESULT 6  
 O98N06 PRELIMINARY; PRT; 155 AA.  
 ID O98N06;  
 AC O98N06;  
 DT 01-OCT-2001 (TRENBLrel. 18, Created)  
 DT 01-OCT-2001 (TRENBLrel. 18, Last sequence update)  
 DT 01-MAR-2002 (TRENBLrel. 20, Last annotation update)  
 DE Hypothetical protein mlr0359.  
 GN MLR0359.  
 OS Rhizobium loti (Mesorhizobium loti).  
 OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;  
 OC Phyllobacteriaceae; Mesorhizobium.  
 CX NCBI\_TaxID=381;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=MAFF303099;  
 RX MEDLINE=21082930; PubMed=11214968;  
 RX Kaneko T., Nakamura Y., Sato S., Asamizu E., Kato T., Sasamoto S.,  
 RA Watanabe A., Idegawa K., Ishikawa A., Kawashima K., Kimura T.,  
 RA Kishida Y., Kiyokawa C., Kohara M., Matsushita M., Matsuno A.,  
 RA Mochizuki Y., Nakayama S., Nakazaki N., Shimo S., Sugimoto M.,  
 RA Takeuchi C., Yamada M., Tabata S.;  
 RT "Complete genome structure of the nitrogen-fixing symbiotic bacterium  
 RT Mesorhizobium loti,"  
 RL DNA Res. 7:331-338(2000).  
 DR EMBL; AP002994; BAB47957.1; -.  
 KW Hypothetical protein; Complete proteome.  
 SQ SEQUENCE 155 AA; 17292 MW; B72785A5A614588 CRC64;

Query Match 37.5%; Score 6; DB 16; Length 155;  
 Best Local Similarity 100.0%; Pred. No. 26;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2 GYGRD 7  
 |||||  
 DB 102 GYGRD 107

## RESULT 7

0826S4 PRELIMINARY; PRT; 160 AA.  
 ID 0826S4  
 AC 0826S4  
 DT 01-MAR-2002 (TRENBLrel. 20, Created)  
 DT 01-MAR-2002 (TRENBLrel. 20, Last sequence update)  
 DT 01-MAR-2002 (TRENBLrel. 20, Last annotation update)  
 DE Alternative bacteriophage tail fiber C-terminus (fragment).  
 GN STY1642.  
 OS Salmonella typhi.  
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;  
 OC Salmonella.  
 OX NCBI\_TaxID=601;  
 OX [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=CT18;  
 RX MEDLINE=21534947; PubMed=11677608;  
 RA Parkhill J., Dougan G., James K.D., Thomson N.R., Pickard D., Wain J.,  
 RA Churcher C., Mungall K.L., Bentley S.D., Holden M.T.G., Sebahia M.,  
 RA Baker S., Basham D., Brooks K., Chillingworth T., Connor P.,  
 RA Cronin A., Davis P., Davies R.M., Dowd L., White N., Farrar J.,  
 RA Feltwell T., Hamlin N., Haque A., Hien T.T., Holroyd S., Jorgels K.,  
 RA Krogan A., Larsen T.S., Leach S., Moule S., O'Gaora P., Parry C.,  
 RA Quail M., Rutherford K., Simmonds M., Skelton J., Stevens K.,  
 RA Whitehead S., Barrell B.G.;  
 RT "Complete genome sequence of a multiple drug resistant Salmonella  
 RT enterica serovar Typhi CT18."  
 RL Nature 413:848-852(2001).  
 DR EMBL; AF627270; CAD01888.1; -.  
 FT Complete proteome.  
 KW NON TER 160  
 SQ SEQUENCE 160 AA; 17368 MW; 5A5B15ADC445E09A CRC64;

Query Match 37.5%; Score 6; DB 16; Length 160;  
 Best Local Similarity 100.0%; Pred. No. 27;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 4 GRDND 9  
 |||||  
 Db 86 GRDND 91

## RESULT 8

092WM3 PRELIMINARY; PRT; 168 AA.  
 ID 092WM3  
 AC 092WM3  
 DT 01-MAY-1999 (TRENBLrel. 10, Created)  
 DT 01-MAY-1999 (TRENBLrel. 10, Last sequence update)  
 DT 01-MAY-1999 (TRENBLrel. 10, Last annotation update)  
 DE Glycine-rich protein-1.  
 OS Cucumis sativus (Cucumber).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;  
 OC eucosids I; Cucurbitales; Cucurbitaceae; Cucumis.  
 OX NCBI\_TaxID=3659;  
 OX [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=ROOT;  
 RC Sakuta C., Satoh S.;  
 RA "Cucumber root glycine-rich protein-1."  
 RL Submitted (JUN-1998) to the EMBL/Genbank/DBJ databases.  
 DR EMBL; AB015173; BAA34063.1; -.  
 SQ SEQUENCE 168 AA; 16800 MW; 62B945E602E5827E CRC64;

Query Match 37.5%; Score 6; DB 10; Length 168;  
 Best Local Similarity 100.0%; Pred. No. 28;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 VGYRD 6  
 |||||  
 Db 53 VGYRD 58

## RESULT 9

09TVR3 PRELIMINARY; PRT; 216 AA.  
 ID 09TVR3  
 AC 09TVR3  
 DT 01-MAY-2000 (TRENBLrel. 13, Created)  
 DT 01-MAY-2000 (TRENBLrel. 13, Last sequence update)  
 DT 01-JUN-2001 (TRENBLrel. 17, Last annotation update)  
 DE Secreted ferritin S subunit precursor.  
 GN FER.  
 OS Calpodas ethiue (Brazilian skipper butterfly).  
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
 OC Pterygota; Neoptera; Endopterygota; Lepidoptera; Glossata; Ditrysia;  
 OC Hesperioidea; Hesperilidae; Hesperilinae; Calpodas.  
 OX NCBI\_TaxID=73634;  
 OX [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=MIDGUT; AND FAT BODY;  
 RX MEDLINE=20023929; PubMed=10560139;  
 RA Nichol H., Locke M.;  
 RT "Secreted ferritin subunits are of two kinds in insects - Molecular  
 RT cloning of cDNAs encoding two major subunits of secreted ferritin from  
 RT Calpodas ethiue."  
 RL Insect Biochem. Mol. Biol. 29:999-1013(1999).  
 DR EMBL; AF161708; AAD50239.1; -.  
 DR HSSP; P02794; 2PHA.  
 DR InterPro: IPR001519; Ferritin.  
 DR Pfam; PF00210; ferritin; 1.  
 DR ProDom; PD000971; Ferritin; 1.  
 KW SIGNAL.  
 FT CHAIN 1 20 POTENTIAL.  
 SQ SEQUENCE 216 AA; 24304 MW; 9A83BF6C05A41CFC CRC64;

Query Match 37.5%; Score 6; DB 5; Length 216;  
 Best Local Similarity 100.0%; Pred. No. 35;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 4 GRDND 9  
 |||||  
 Db 158 GRDND 163

## RESULT 10

P95893 PRELIMINARY; PRT; 243 AA.  
 ID P95893  
 AC P95893  
 DT 01-MAY-1997 (TRENBLrel. 03, Created)  
 DT 01-MAY-1997 (TRENBLrel. 03, Last sequence update)  
 DT 01-MAR-2002 (TRENBLrel. 20, Last annotation update)  
 DE Hypothetical protein SSO2119.  
 GN SSO2119 OR CO2013.  
 OS Sulfolobus solfataricus.  
 OC Archaea; Crenarchaeota; Thermoprotei; Sulfolobales; Sulfolobaceae;  
 OC Sulfolobus.  
 OX NCBI\_TaxID=2287;  
 OX [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=ATCC 35092 / DSM 1617 / P2;  
 RX MEDLINE=2133296; PubMed=11427726;  
 RA She Q., Singh R.K., Confalonieri F., Zivanovic Y., Allard G.,  
 RA Aveyez M.J., Chan-Weher C.C.-Y., Clausen I.G., Curtis B.A.,  
 RA De Moers A., Eranuo G., Fletcher C., Gordon P.M.K.,  
 RA Heikamp-de Jong I., Jeffries A.C., Kozera C.J., Medina N., Peng X.,  
 RA Thi-Ngoc H.P., Redder P., Schenk M.E., Theriault C., Tolstrup N.,  
 RA Charlebois R.L., Doolittle W.F., Duguet M., Gaasterland T.,  
 RA Garrett R.A., Regan M.A., Sensen C.W., Van der Oost J.;  
 RT "The complete genome of the crenarchaeon Sulfolobus solfataricus P2."  
 RL Proc. Natl. Acad. Sci. U.S.A. 98:7835-7840(2001).  
 DR EMBL; Y08256; CAA69445.1; -.  
 DR EMBL; AE006818; AAK42298.1; -.  
 DR InterPro; IPR002931; Trnsglutase\_1like.

DR Pfam; PF01841; Transglut\_core; 1.  
 DR SMART; SM00460; TGC; 1.  
 KW Hypothetical protein; Complete proteome.  
 SQ SEQUENCE 243 AA; 28157 MW; 7846D2E5728B4D2 CRC64;

Query Match 37.5%; Score 6; DB 17; Length 243;  
 Best Local Similarity 100.0%; Pred. No. 38;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4 GYGRDD 9  
 Db 216 GYGRDD 221

RESULT 11  
 Q8U9E7 PRELIMINARY; PRT; 244 AA.  
 AC Q8U9E7;  
 DT 01-JUN-2002 (TRENBLrel. 21, Created)  
 DT 01-JUN-2002 (TRENBLrel. 21, Last sequence update)  
 DT 01-JUN-2002 (TRENBLrel. 21, Last annotation update)  
 DE Two component response regulator.  
 GN ATU3781 OR AGR\_L\_2108.  
 OS Agrobacterium tumefaciens (strain C58 / ATCC 33970).  
 OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;  
 OC Rhizobiaceae; Rhizobium.  
 OX NCBI\_TaxID=176299;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=21608550; PubMed=11743193;  
 RA Wood D.W., Setubal J.C., Kaul R., Monks D.E., Kitajima J.P.,  
 RA Okura Y.K., Zhou Y., Chen L., Wood G.E., Almeida N.F. Jr., Woo L.,  
 RA Chen Y., Paulsen I.T., Eisen J.A., Karp P.D., Boye D. Sr.,  
 RA Chapman P., Clendinning J., Deatherage G., Gillet W., Grant C.,  
 RA Kutayavin T., Levy R., Li M.-J., McClelland E., Palmeri A.,  
 RA Raymond C., Rouse G., Saenphimmachak C., Wu Z., Romero P., Gordon D.,  
 RA Zhang S., Yoo H., Tao Y., Biddle P., Jung M., Krespan W., Perry M.,  
 RA Gordon-Kamm B., Liao L., Kim S., Hendrick C., Zhao Z.-Y., Dolan M.,  
 RA Chumley F., Tingey S.V., Tomb J.-F., Gordon M.P., Olson M.V.,  
 RA Nester E.W.;  
 RT "The genome of the natural genetic engineer Agrobacterium tumefaciens C58.";  
 RL Science 294:2317-2323 (2001).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=21608551; PubMed=11743194;  
 RA Goodner B., Hinkle G., Gattung S., Miller N., Blanchard M.,  
 RA Qurollo B., Goldman B.S., Cao Y., Askenazi M., Halling C., Mullin L.,  
 RA Houmlel K., Gordon J., Vaudin M., Iartchouk O., Epp A., Liu F.,  
 RA Wollam C., Allinger M., Doughy D., Scott C., Lappas C., Markelz B.,  
 RA Flanagan C., Crowell C., Gursen J., Lomo C., Sear C., Strub G.,  
 RA Cialo C., Slater S.;  
 RT "Genome sequence of the plant pathogen and biotechnology agent Agrobacterium tumefaciens C58.";  
 RL Science 294:2323-2328 (2001).  
 DR EMBL; AE009309; AAL44591.1; ALT\_INIT.  
 DR EMBL; AE008305; AAK89627.1; -.  
 KW Complete proteome.  
 SQ SEQUENCE 244 AA; 27202 MW; 8FDFAA807681CB5B CRC64;

Query Match 37.5%; Score 6; DB 16; Length 244;  
 Best Local Similarity 100.0%; Pred. No. 39;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 GYGRDD 7  
 Db 14 GYGRDD 19

RESULT 12  
 Q98HG8 PRELIMINARY; PRT; 294 AA.  
 ID Q98HG8;  
 AC Q98HG8;

DT 01-OCT-2001 (TRENBLrel. 18, Created)  
 DT 01-OCT-2001 (TRENBLrel. 18, Last sequence update)  
 DT 01-MAR-2002 (TRENBLrel. 20, Last annotation update)  
 DE Hypothetical protein ml12875.  
 GN ML12875.  
 OS Rhizobium loti (Mesorhizobium loti).  
 OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;  
 OC Phyllobacteriaceae; Mesorhizobium.  
 OX NCBI\_TaxID=381;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX STRAIN=MAFF303099; PubMed=11214968;  
 RA Kaneko T., Nakamura Y., Sato S., Asamizu E., Kato T., Sasamoto S.,  
 RA Watanabe A., Idegawa K., Ishikawa A., Kawashina K., Kimura T.,  
 RA Kishida Y., Kiyokawa C., Kohara M., Matsumoto M., Matsuno A.,  
 RA Mochizuki Y., Nakayama S., Nakazaki N., Shimizu S., Sugimoto M.,  
 RA Takeuchi C., Yamada M., Tabata S.;  
 RT "Complete genome structure of the nitrogen-fixing symbiotic bacterium Mesorhizobium loti.";  
 RL DNA Res. 7:331-338 (2000).  
 DR EMBL; AP003000; BAB49898.1; -.  
 DR InterPro; IPR02931; Trnsglutase\_1like.  
 DR Pfam; PF01841; Transglut\_core; 1.  
 DR SMART; SM00460; TGC; 1.  
 KW Hypothetical protein; Complete proteome.  
 SQ SEQUENCE 294 AA; 31526 MW; 0BB774741F23E29B CRC64;

Query Match 37.5%; Score 6; DB 16; Length 294;  
 Best Local Similarity 100.0%; Pred. No. 45;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VGYGRD 6  
 Db 256 VGYGRD 261

RESULT 13  
 Q8U9E2 PRELIMINARY; PRT; 303 AA.  
 AC Q8U9E2;  
 DT 01-JUN-2002 (TRENBLrel. 21, Created)  
 DT 01-JUN-2002 (TRENBLrel. 21, Last sequence update)  
 DT 01-JUN-2002 (TRENBLrel. 21, Last annotation update)  
 DE Hypothetical protein Atu3867.  
 GN ATU3867 OR AGR\_L\_1957.  
 OS Agrobacterium tumefaciens (strain C58 / ATCC 33970).  
 OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;  
 OC Rhizobiaceae; Rhizobium.  
 OX NCBI\_TaxID=176299;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=21608550; PubMed=11743193;  
 RA Wood D.W., Setubal J.C., Kaul R., Monks D.E., Kitajima J.P.,  
 RA Okura Y.K., Zhou Y., Chen L., Wood G.E., Almeida N.F. Jr., Woo L.,  
 RA Chen Y., Paulsen I.T., Eisen J.A., Karp P.D., Boye D. Sr.,  
 RA Chapman P., Clendinning J., Deatherage G., Gillet W., Grant C.,  
 RA Kutayavin T., Levy R., Li M.-J., McClelland E., Palmeri A.,  
 RA Raymond C., Rouse G., Saenphimmachak C., Wu Z., Romero P., Gordon D.,  
 RA Zhang S., Yoo H., Tao Y., Biddle P., Jung M., Krespan W., Perry M.,  
 RA Gordon-Kamm B., Liao L., Kim S., Hendrick C., Zhao Z.-Y., Dolan M.,  
 RA Chumley F., Tingey S.V., Tomb J.-F., Gordon M.P., Olson M.V.,  
 RA Nester E.W.;  
 RT "The genome of the natural genetic engineer Agrobacterium tumefaciens C58.";  
 RL Science 294:2317-2323 (2001).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=21608551; PubMed=11743194;  
 RA Goodner B., Hinkle G., Gattung S., Miller N., Blanchard M.,  
 RA Qurollo B., Goldman B.S., Cao Y., Askenazi M., Halling C., Mullin L.,  
 RA Houmlel K., Gordon J., Vaudin M., Iartchouk O., Epp A., Liu F.,  
 RA Wollam C., Allinger M., Doughy D., Scott C., Lappas C., Markelz B.,

RA Flanagan C., Crowell C., Garson J., Lomo C., Sear C., Strub G.,  
RA Cleio C., Slater S.;  
RT "Genome sequence of the plant pathogen and biotechnology agent  
RT Agrobacterium tumefaciens C58.";  
RL Science 294:2323-2328(2001).  
DR EMBL; AE009337; AAL44676.1; ALT\_INIT.  
DR EMBL; AE008297; AAK89552.1; -.  
RW Hypothetical protein; Complete proteome.  
SQ SEQUENCE 303 AA; 33374 MW; 92F95B275A34FE17 CRC64;

Query Match 37.5%; Score 6; DB 16; Length 303;  
Best Local Similarity 100.0%; Pred. No. 47;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 VGYGRD 6  
Db 269 VGYGRD 274

## RESULT 14

Q8VG21 PRELIMINARY; PRT; 308 AA.  
ID Q8VG21;  
AC Q8VG21;  
DT 01-MAR-2002 (TREMBLrel. 20, Created)  
DT 01-MAR-2002 (TREMBLrel. 20, Last sequence update)  
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)  
DE Olfactory receptor MOR230-4.  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sclurognathi; Muridae; Murinae; Mus.  
CX NCBI\_TaxID=10090;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Zhang X., Firestein S.J.;  
RT "The olfactory receptor gene superfamily of the mouse."  
RL Nat. Neurosci. 0:0-0(2002).  
RN [2]  
RP SEQUENCE FROM N.A.  
RA Adams M.;  
RL Submitted (JAN-2002) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AY073342; AAL61005.1; -.  
DR InterPro; IPR000276; GPCR\_Rhodpsn.  
DR Pfam; PF00001; 7tm\_1; 1.  
DR PRINTS; PR00237; GPCR\_RHODOPSN.  
DR PROSITE; PS00237; G\_PROTEIN\_RECPE\_F1\_1; UNKNOWN\_1.  
DR PROSITE; PS50262; G\_PROTEIN\_RECPE\_F1\_2; 1.  
KW Receptor.

SQ SEQUENCE 308 AA; 34854 MW; 332E63DF0ECBC088 CRC64;

Query Match 37.5%; Score 6; DB 11; Length 308;  
Best Local Similarity 100.0%; Pred. No. 47;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 11 DPSRKN 16  
Db 18 DPSRKN 23

## RESULT 15

Q8VF94 PRELIMINARY; PRT; 310 AA.  
ID Q8VF94;  
AC Q8VF94;  
DT 01-MAR-2002 (TREMBLrel. 20, Created)  
DT 01-MAR-2002 (TREMBLrel. 20, Last sequence update)  
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)  
DE Olfactory receptor MOR230-8.  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sclurognathi; Muridae; Murinae; Mus.  
CX NCBI\_TaxID=10090;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Zhang X., Firestein S.J.;

Qy 11 DPSRKN 16  
Db 18 DPSRKN 23

RT "The olfactory receptor gene superfamily of the mouse."  
RL Nat. Neurosci. 0:0-0(2002).  
RN [2]  
RP SEQUENCE FROM N.A.  
RA Adams M.;  
RL Submitted (JAN-2002) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AY073640; AAL61303.1; -.  
DR InterPro; IPR000276; GPCR\_Rhodpsn.  
DR Pfam; PF00001; 7tm\_1; 1.  
DR PRINTS; PR00237; GPCR\_RHODOPSN.  
DR PROSITE; PS00237; G\_PROTEIN\_RECPE\_F1\_1; UNKNOWN\_1.  
DR PROSITE; PS50262; G\_PROTEIN\_RECPE\_F1\_2; 1.  
KW Receptor.

SQ SEQUENCE 310 AA; 34981 MW; C392CD833D39EA7 CRC64;

Query Match 37.5%; Score 6; DB 11; Length 310;  
Best Local Similarity 100.0%; Pred. No. 48;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 11 DPSRKN 16  
Db 18 DPSRKN 23

## RESULT 16

Q8YS18 PRELIMINARY; PRT; 335 AA.  
ID Q8YS18;  
AC Q8YS18;  
DT 01-MAR-2002 (TREMBLrel. 20, Created)  
DT 01-MAR-2002 (TREMBLrel. 20, Last sequence update)  
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)  
DE Hypothetical protein ALR3097.  
GN ALR3097.  
OS Anabaena sp. (strain PCC 7120).  
OC Bacteria; Cyanobacteria; Nostocales; Nostocaceae; Nostoc.  
CX NCBI\_TaxID=103690;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=21595285; PubMed=11759840;  
RA Kaneko T., Nakamura Y., Wolk C.P., Kuritz T., Sasamoto S.,  
RA Kawanabe A., Iriguchi M., Ishikawa A., Kawashima K., Kimura T.,  
RA Kishida Y., Kohara M., Matsumoto M., Matsuno A., Muraki A.,  
RA Nakazaki N., Shimoto S., Sugimoto M., Takazawa M., Yamada M.,  
RA Yaouda M., Tabata S.;  
RT "Complete genomic sequence of the filamentous nitrogen-fixing  
RT cyanobacterium Anabaena sp. strain PCC 7120."  
RL DNA Res. 8:205-213(2001).  
DR EMBL; AP003591; BAB74796.1; -.  
DR InterPro; IPR000345; CytoC\_heme\_bind.  
DR PROSITE; PS00190; CYTOCHROME\_C; UNKNOWN\_1.  
KW Hypothetical protein; Complete proteome.

SQ SEQUENCE 335 AA; 37716 MW; 5E22200A4F643477 CRC64;

Query Match 37.5%; Score 6; DB 16; Length 335;  
Best Local Similarity 100.0%; Pred. No. 51;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 VGYGRD 6  
Db 169 VGYGRD 174

## RESULT 17

Q9KK91 PRELIMINARY; PRT; 366 AA.  
ID Q9KK91;  
AC Q9KK91;  
DT 01-OCT-2000 (TREMBLrel. 15, Created)  
DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)  
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)  
DE Chitopodin.  
GN CHIP.  
OS Vibrio furnissii.  
OC Bacteria; Proteobacteria; gamma subdivision; Vibrionaceae; Vibrio.

Qy 1 VGYGRD 6  
Db 169 VGYGRD 174

```

CX NCBI_TaxID=29494;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20490750; PubMed=10913115;
RA Keyhani N.O., Li X.B., Roseman S.;
RT "Chitin catabolism in the marine bacterium Vibrio furnissii."
RT J. Biol. Chem. 275:33068-33076(2000).
RL Identificaiton and molecular cloning of a chitinoritin.";
DR EMBL; AF129934; AAF97616.1; -.
DR InterPro; IPR001702; Protin_glam-ve.
DR Pfam; PF00267; Gram-ve_porins; 1.
DR PRINTS; PR00182; ECOLNEIPORIN.
KW Porin.
SQ SEQUENCE 366 AA; 39310 MM; 35BB8C6BE27090E7 CRC64;

Query Match 37.5%; Score 6; DB 2; Length 366;
Best Local Similarity 100.0%; Pred.No.55;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4 GRDND 9
    |||||
Db 180 GRDND 185

RESULT 18
Q9PRG7 PRELIMINARY; PRT; 401 AA.
ID Q9PRG7
AC Q9PRG7;
DT 01-MAY-2000 (TRENBLREL. 13, Created)
DT 01-MAY-2000 (TRENBLREL. 13, Last sequence update)
DT 01-JUN-2001 (TRENBLREL. 17, Last annotation update)
DE P75 neurotrophin receptor A-2.
GN P75NTRA.
OS Xenopus laevis (African clawed frog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipidea; Pipidae;
OC Xenopodidae; Xenopus.
CX NCBI_TaxID=8355;
RN [1]
RP SEQUENCE FROM N.A.
RA Huxson L.D., Richards A.P., Bothwell M.;
RT "Life and death in the developing Xenopus retina: The role of the p75
RT neurotrophin receptor.";
RL Submitted (JUL-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF172400; AAD51031.1; -.
DR EMBL; AF172399; AAD51030.1; -.
DR HSP; P07174; INGR.
DR InterPro; IPR000488; Death.
DR InterPro; IPR000561; EGF-like.
DR InterPro; IPR001368; TNFR_c6.
DR Pfam; PF00531; death; 1.
DR Pfam; PF00020; TNFR_c6; 4.
DR SMART; SM00005; DEATH; 1.
DR SMART; SM00208; TNFR; 4.
DR PROSITE; PS0017; DEATH_DOMAIN; 1.
DR PROSITE; PS01186; EGF_2; UNKNOWN_1.
DR PROSITE; PS00652; TNFR_NGFR_1; UNKNOWN_3.
DR PROSITE; PS00650; TNFR_NGFR_2; 3.
DR PROSITE; PS00650; TNFR_NGFR_2; 3.
KW Receptor.
SQ SEQUENCE 401 AA; 43419 MM; D13BCAF7863EECF CRC64;

Query Match 37.5%; Score 6; DB 13; Length 401;
Best Local Similarity 100.0%; Pred.No.60;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 7 DNRDP 12
    |||||
Db 148 DNRDP 153

RESULT 19
053655 PRELIMINARY; PRT; 412 AA.
ID 053655
```

```

AC 053655;
DT 01-JUN-1998 (TRENBLREL. 06, Created)
DT 01-JUN-1998 (TRENBLREL. 06, Last sequence update)
DT 01-JUN-2002 (TRENBLREL. 21, Last annotation update)
DE Hypothetical protein Rv0204c.
GN Rv0204c OR MT0214 OR MT0033.12C.
OS Mycobacterium tuberculosis.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
CX NCBI_TaxID=1773;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=H37RV;
RX MEDLINE=98295987; PubMed=9634230;
RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
RA Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekala F.,
RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA Davies R., Devlin K., Felwell T., Gentles S., Hamlin N., Holroyd S.,
RA Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
RA Rutter S., Seeger K., Skelton S., Squares S., Squares R.,
RA Sulston J.E., Taylor K., Whitehead S., Barrell B.G.;
RT "Deciphering the biology of Mycobacterium tuberculosis from the
RT complete genome sequence.";
RL Nature 393:537-544(1998).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=COC 1551 / OSHKOSH;
RA Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,
RA Peterson J., Deboy R., Dodson R., Gwinn M., Haft D., Hickey E.,
RA Kolony J.F., Nelson W.C., Umayam L.A., Ermolaeva M., Salzberg S.L.,
RA Delcher A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A.,
RA Bishai W.;
RT "Whole genome comparison of Mycobacterium tuberculosis clinical and
RT laboratory strains.";
RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AL021928; CAA17326.1; -.
DR EMBL; AE006931; AAK4435.1; -.
DR TIGR; MT0214; -.
DR TubercuList; Rv0204c; -.
DR InterPro; IPR005242; Cons_hypoch374.
DR Pfam; PF03706; UPF0104; 1.
DR TIGRFAMs; TIGR00374; TIGR00374; 1.
KW Hypothetical protein; Complete proteome.
FT CONFLICT 306 V -> L (IN REF. 2).
SQ SEQUENCE 412 AA; 44022 MM; C24A19AFC1732C90 CRC64;

Query Match 37.5%; Score 6; DB 16; Length 412;
Best Local Similarity 100.0%; Pred.No.61;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 7 DNRDP 12
    |||||
Db 363 DNRDP 368

RESULT 20
Q9DCV1 PRELIMINARY; PRT; 461 AA.
ID Q9DCV1
AC Q9DCV1;
DT 01-JUN-2001 (TRENBLREL. 17, Created)
DT 01-JUN-2001 (TRENBLREL. 17, Last sequence update)
DT 01-JUN-2002 (TRENBLREL. 21, Last annotation update)
DE Adult male kidney cDNA, RIKEN full-length enriched library,
DE clone:0610010C21, full insert sequence.
GN CTSC.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
CX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=KIDNEY;
```

RX MEDLINE=21085660; PubMed=11217851;  
 RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,  
 RA Aikawa T., Hara A., Fukunishi Y., Kono H., Adachi J., Fukuda S.,  
 RA Aikawa K., Iwata M., Nishi K., Kiyosawa H., Kondo S., Yamanka I.,  
 RA Saito T., Okazaki Y., Gotohori T., Bono H., Kaikawa T., Saito R.,  
 RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,  
 RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochwa H.,  
 RA Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,  
 RA Schiraldi L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,  
 RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,  
 RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,  
 RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,  
 RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamuya M., Lee N.H.,  
 RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombarts P.,  
 RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,  
 RA Sasaki H., Sato K., Schoenbach C., Seva T., Shibata Y., Storch K.-P.,  
 RA Suzuki H., Toyooka K., Wang K.H., Weitz C., Whitaker C., Wilming L.,  
 RA Wyszewski-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohsaki S.,  
 RA Hayashizaki Y.,  
 RT "Functional annotation of a full-length mouse cDNA collection."  
 RL Nature 409:685-690(2001).  
 DR EMBL; AK002454; BAB22112.1; -  
 DR HSSP; P07688; 1DDO.  
 DR MEROPS; C01.070; -;  
 DR MGD; MGI:109553; Ctsc.  
 DR InterPro; IPR000668; Peptidase\_C1.  
 DR InterPro; IPR000169; SHPrc acsite.  
 DR Pfam; PF00112; Peptidase\_C1; 1.  
 DR PRINTS; PR00705; PAPAIN.  
 DR Prodom; PD000156; Peptidase\_C1; 1.  
 DR PROSITE; PS00640; THIOL\_PROTEASE\_ASN; 1.  
 DR PROSITE; PS00139; THIOL\_PROTEASE\_CYS; 1.  
 DR PROSITE; PS00639; THIOL\_PROTEASE\_HIS; 1.  
 SQ SEQUENCE 461 AA; 52392 MW; 551E2D35E46ABEB2 CRC64;

Query Match 37.5%; Score 6; DB 11; Length 461;  
 Best Local Similarity 100.0%; Pred. No. 68;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 VGYGRD 6  
 DB 408 VGYGRD 413  
 RESULT 21  
 ID 09Z190 PRELIMINARY; PRT; 478 AA.  
 AC 09Z190;  
 DT 01-MAY-1999 (TREMBlrel. 10, Created)  
 DT 01-MAY-1999 (TREMBlrel. 10, Last sequence update)  
 DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)  
 DE Putative histidine kinase.  
 GN Hpk48.  
 OS Lactobacillus sakei.  
 OC Bacteria; Firmicutes; Bacillus/Clostridium group; Lactobacillales;  
 OC Lactobacillaceae; Lactobacillus.  
 OX NCBI\_TaxID=1599;  
 RX [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=23k;  
 RA Morel-deville F., Fauvel F., Morel P.;  
 RL Submitted (DEC-1997) to the EMBL/GenBank/DBJ databases.  
 CC -1- SIMILARITY TO OTHER PROKARYOTIC SENSORY TRANSDUCTION HISTIDINE KINASES.  
 CC EMBL; AF036968; AAD10268.1; -;  
 DR InterPro; IPR003594; ATPbind\_ATPase.  
 DR InterPro; IPR004358; Bact\_sens\_pr\_C.  
 DR InterPro; IPR003660; HAMP.  
 DR InterPro; IPR003661; His\_kinA.  
 DR InterPro; IPR004359; His\_kin\_51g.  
 DR Pfam; PF00672; HAMP; 1.  
 DR Pfam; PF02518; HATPase\_C; 1.  
 DR Pfam; PF00512; signal; 1.

DR PRINTS; PR00344; BCTR1SENSOR.  
 DR SMART; SM00304; HAMP; 1.  
 DR SMART; SM00387; HATPase\_C; 1.  
 DR SMART; SM00388; HisKA; 1.  
 KW Kinase; Phosphorylation; Sensory transduction; Transferase.  
 SQ SEQUENCE 478 AA; 54402 MW; D79C1704B40849BD CRC64;

Query Match 37.5%; Score 6; DB 2; Length 478;  
 Best Local Similarity 100.0%; Pred. No. 70;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 11 DPSRKN 16  
 DB 423 DPSRKN 428

RESULT 22  
 ID 09G329 PRELIMINARY; PRT; 501 AA.  
 AC 09G329;  
 DT 01-MAR-2001 (TREMBlrel. 16, Created)  
 DT 01-MAR-2001 (TREMBlrel. 16, Last sequence update)  
 DT 01-MAR-2002 (TREMBlrel. 20, Last annotation update)  
 DE Probable intron maturase (Maturase K) (Fragment).  
 GN MATK.  
 OS Vallisneria spiralis.  
 OC Chloroplast.  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; Liliopsida; Hydrocharitaceae;  
 OC Vallisneria.  
 OX NCBI\_TaxID=62345;  
 RX [1]  
 RP SEQUENCE FROM N.A.  
 RA Tanaka N., Setoguchi H., Murata J.;  
 RT "Phylogeny of the Family Hydrocharitaceae Inferred from rbcL and matK  
 RT Gene Sequence Data."  
 RL J. Plant Res. 110:329-337(1997).  
 CC -1- FUNCTION: PROBABLY ASSISTS IN SPLICING CHLOROPLAST GROUP II  
 CC INTRONS (BY SIMILARITY).  
 CC -1- SIMILARITY: WITH CORRESPONDING ORF IN OTHER PLANT CHLOROPLASTS,  
 CC AND REGIONS OF SIMILARITY TO MATURASE-LIKE POLYPEPTIDES ENCODED BY  
 CC MITOCHONDRIAL INTRONS.  
 CC EMBL; AB002578; BAB18668.1; -;  
 DR InterPro; IPR000442; Intron\_mature2.  
 DR InterPro; IPR002866; MATK\_N.  
 DR Pfam; PF01348; Intron\_mature2; 1.  
 DR Pfam; PF01824; MATK\_N; 1.  
 KW Chloroplast; mRNA processing.  
 FT NON\_TER 501 501  
 SQ SEQUENCE 501 AA; 60407 MW; 424FAF8933F6224 CRC64;

Query Match 37.5%; Score 6; DB 8; Length 501;  
 Best Local Similarity 100.0%; Pred. No. 73;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2 GYGRDD 7  
 DB 164 GYGRDD 169

RESULT 23  
 ID P72957 PRELIMINARY; PRT; 509 AA.  
 AC P72957;  
 DT 01-FEB-1997 (TREMBlrel. 02, Created)  
 DT 01-FEB-1997 (TREMBlrel. 02, Last sequence update)  
 DT 01-MAR-2002 (TREMBlrel. 20, Last annotation update)  
 DE Hypothetical protein sil10641.  
 GN SL10641.  
 OS Synecocystis sp. (strain PCC 6803).  
 OC Bacteria; Cyanobacteria; Chroococcales; Synecocystis.  
 OX NCBI\_TaxID=1148;  
 RX [1]



RP SEQUENCE FROM N.A.  
 RX MEDLINE=97061201; PubMed=89905231;  
 RA Kaneko T., Sato S., Kotani H., Tanaka A., Asamizu E., Nakamura Y.,  
 RA Miyajima N., Hirosewa M., Sugita M., Sasamoto S., Kimura T.,  
 RA Hosouchi T., Matsumoto A., Muraki A., Nakazaki N., Naruo K., Okumura S.,  
 RA Shimpou S., Takeuchi C., Wada T., Watanabe A., Yamada M., Yasuda M.,  
 RA Tabata S.;  
 RT "Sequence analysis of the genome of the unicellular cyanobacterium  
 RT *Synechocystis* sp. strain PCC6803. II. Sequence determination of the  
 RT entire genome and assignment of potential protein-coding regions.";  
 RL DNA Res.3:109-136(1996).  
 DR EMBL; D90902; BAA16975.1; -;  
 KW Hypothetical protein; Complete proteome.  
 SQ SEQUENCE 509 AA; 58509 MW; 100354B5139315BC CRC64;

Query March 37.5%; Score 6; DB 16; Length 509;  
 Best Local Similarity 100.0%; Pred. No. 74;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 VQYGRD 6  
 |||||  
 Db 376 VQYGRD 381

RESULT 24  
 Q8T1N3 PRELIMINARY; PRT; 515 AA.  
 AC Q8T1N3;  
 DT 01-JUN-2002 (TREMBlrel. 21, Created)  
 DT 01-JUN-2002 (TREMBlrel. 21, Last sequence update)  
 DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)  
 DE Hypothetical 61.5 kDa protein.  
 OS Dictyostelium discoideum (Slime mold).  
 OC Eukaryota; Mycetozoa; Dictyostelidae; Dictyostelium.  
 OX NCBI\_TaxID=44689;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STOECHN=AK4;  
 RA Gloechner G., Eichinger L., Szafranski K., Pachebat J., Dear P.,  
 RA Lehmann R., Baugart C., Parra G., April J.F., Guigo R., Kumpf K.,  
 RA Tungsall B., Cox E., Quail M.A., Platzer M., Rosenthal A., Noegel A.A.;  
 RT "Sequence and Analysis of Chromosome 2 of Dictyostelium";  
 RL Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AC116330; AAL96720.1; -;  
 KW Hypothetical protein.  
 SQ SEQUENCE 515 AA; 61503 MW; 2761DE2105859E9F CRC64;

Query March 37.5%; Score 6; DB 5; Length 515;  
 Best Local Similarity 100.0%; Pred. No. 74;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 6 DDNDRD 11  
 |||||  
 Db 331 DDNDRD 336

RESULT 25  
 Q8T4G4 PRELIMINARY; PRT; 774 AA.  
 AC Q8T4G4;  
 DT 01-JUN-2002 (TREMBlrel. 21, Created)  
 DT 01-JUN-2002 (TREMBlrel. 21, Last sequence update)  
 DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)  
 DE SPO1877P.  
 OS Drosophila melanogaster (Fruit fly).  
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
 OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
 OC Ephydroidea; Drosophilidae; Drosophila.  
 OX NCBI\_TaxID=7227;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Stapleton M., Brokstein P., Hong L., Agdayani A., Carlson J.,

RA Champe M., Chavez C., Dorsett V., Dresnek D., Farfan D., Frise E.,  
 RA George R., Gonzalez M., Guain H., Krommiller B., Li P., Liao G.,  
 RA Miranda A., Murgall C.J., Nunoo J., Pacble V., Paragas V., Park S.,  
 RA Patel S., Phuanenavong S., Wan K., Yu C., Lewis S.E., Rubin G.M.,  
 RA Seinkner S.;  
 RL Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AY084200; AAL89938.1; -;  
 KW SEQUENCE 774 AA; 85959 MW; F613E3F7A11A75DC CRC64;

Query March 37.5%; Score 6; DB 5; Length 774;  
 Best Local Similarity 100.0%; Pred. No. 11e+02;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 9 DRDPGR 14  
 |||||  
 Db 546 DRDPGR 551

RESULT 26  
 Q9C5J3 PRELIMINARY; PRT; 946 AA.  
 AC Q9C5J3;  
 DT 01-JUN-2001 (TREMBlrel. 17, Created)  
 DT 01-JUN-2001 (TREMBlrel. 17, Last sequence update)  
 DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)  
 DE Hypothetical 108.3 kDa protein.  
 GN T1H3 70/AT5G25060.  
 OS Arabidopsis thaliana (Mouse-ear cress).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;  
 OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.  
 OX NCBI\_TaxID=3702;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Yamada K., Liu S.X., Sakano H., Pham P.K., Banh J., Chung M.K.,  
 RA Goldsmith A.D., Lee J.M., Quach H.L., Toriumi M., Yu G., Bowser L.,  
 RA Carninci P., Chen H., Cheuk R., Hayashizaki Y., Ishida J., Jones T.,  
 RA Kamita A., Karlin-Neumann G., Kawai J., Kim C., Lam B., Lin J.,  
 RA Miranda M., Narusaka M., Nguyen M., Palm C.J., Sakurai T., Satou M.,  
 RA Seki M., Shim P., Southwick A., Shinzaki K., Davis R.W., Ecker J.R.,  
 RA Theologis A.;  
 RT "Full Length cDNA of gene T1H3 70/AT5G25060.";  
 RL Submitted (MAR-2001) to the EMBL/GenBank/DBJ databases.  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RA Yamada K., Liu S.X., Sakano H., Pham P.K., Banh J., Chung M.K.,  
 RA Dale J.M., Gibson H.A., Goldsmith A.D., Jiang P.K., Lee J.M.,  
 RA Quach H.L., Tang C., Toriumi M., Yu G., Bowser L., Carninci P.,  
 RA Chen H., Cheuk R., Hayashizaki Y., Ishida J., Jones T., Kamita A.,  
 RA Karlin-Neumann G., Kawai J., Kim C., Koesema E., Lam B., Lin J.,  
 RA Meyers M.C., Miranda M., Narusaka M., Nguyen M., Palm C.J.,  
 RA Sakurai T., Satou M., Seki M., Shim P., Southwick A., Tracy S.E.,  
 RA Shinzaki K., Davis R.W., Ecker J.R., Theologis A.;  
 RT "Full length cDNA of gene T1H3 70/AT5G25060.";  
 RL Submitted (JUN-2001) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AF360212; AAK25922.1; -;  
 DR EMBL; AY040060; AAK64118.1; -;  
 DR HSSP; P11940; ICVJ.  
 DR InterPro; IPR000504; RNA\_rec\_mot.  
 DR InterPro; IPR003034; SAP.  
 DR InterPro; IPR000061; Surp.  
 DR Pfam; PF00076; rrm; 1.  
 DR Pfam; PF01805; Surp; 1.  
 DR SMART; SM00360; RRM; 1.  
 DR PROSITE; PS50102; RRM; 1.  
 KW Hypothetical protein.  
 SQ SEQUENCE 946 AA; 108285 MW; B531054388A8EB8B CRC64;

Query March 37.5%; Score 6; DB 10; Length 946;  
 Best Local Similarity 100.0%; Pred. No. 1.3e+02;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 5 RDDNDR 10

Db 927 RDDNDR 932

## RESULT 27

ID Q9Y142 PRELIMINARY; PRT; 985 AA.

DT 01-NOV-1999 (TRENBLREL. 12, Created)  
DT 01-NOV-1999 (TRENBLREL. 12, Last sequence update)  
DT 01-MAR-2001 (TRENBLREL. 16, Last annotation update)

DE BCNA:GH05582

GN BCNA:GH05582 OR CG8740.

OS Drosophila melanogaster (Fruit fly).

OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;

OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;

OC Ephydroidea; Drosophilidae; Drosophila.

OX NCBI\_TaxID=7227;

RN [1]

RP SEQUENCE FROM N.A.

RA Rubin G.M., Wan K.H., Harvey D., Lewis S.E., Brokstein P., Tsang G.,

RA Agbayani A., Arcina T.T., Baxter E., Blazej R.G., Burenhoff C.,

RA Champe M., Chavez C., Chew M., Doyle C.M., Farfan D.E., Frise E.,

RA Galle R., George R.A., Harris N.L., Hoskins R.A., Evans-Holm M.,

RA Houston K.A., Hummasti S.R., Kim E., Li P., Moshrefi M., Pacleb J.M.,

RA Park S., Sequeira A., Sethi H., Snir E., Svirskaas R.R., Weinburg T.,

RA Ceiniker S.E.;

RT "Full length Drosophila melanogaster cDNA sequence.";

RL Submitted (APR-1999) to the EMBL/GenBank/DBJ databases.

DR EMBL; AF154629; AAD38604.1.

DR FLYBase; FBgn027585; BCNA:GH05582.

SO SEQUENCE 985 AA; 112852 MW; 6687661C0506BFAF CRC64;

Query Match 37.5%; Score 6; DB 5; Length 985;  
Best Local Similarity 100.0%; Pred. No. 1.3e+02;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 VGYGRD 6  
Db 819 VGYGRD 824

## RESULT 28

ID Q9V4X8 PRELIMINARY; PRT; 985 AA.

DT 01-MAY-2000 (TRENBLREL. 13, Created)

DT 01-MAY-2000 (TRENBLREL. 13, Last sequence update)

DT 01-MAR-2001 (TRENBLREL. 16, Last annotation update)

DE BCNA:GH05582 protein.

GN BCNA:GH05582 OR CG8740.

OS Drosophila melanogaster (Fruit fly).

OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;

OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;

OC Ephydroidea; Drosophilidae; Drosophila.

OX NCBI\_TaxID=7227;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=BERKELEY;

DE MEDLINE=20196006; PubMed=10731132;

RA Adams M.D., Ceiniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,

RA Geamatidis P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,

RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,

RA Suton G.G., Wortman J.R., Yandell M.D., Zhang O., Chen L.X.,

RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,

RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,

RA Abtli J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,

RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasly E.M.,

RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,

RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Broctier P.,

RA Butts K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,

RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,

RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,

RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,

RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,

RA Foster C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glasser K.,

RA Glodex A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,

RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,

RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,

RA Jatali M., Kalush F., Karpen G.H., Ke Z., Kenison J.A., Ketchum K.A.,

RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,

RA Laeko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,

RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,

RA Merkulov G., Mishina N.V., Mobarry C., Morris J., Moshrefi A.,

RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,

RA Nelson D.R., Nelson K.A., Nixon K., Nusken D.R., Pacleb J.M.,

RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,

RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,

RA Shue B.C., Siden-Klamos I., Simpson M., Skupski M.P., Smith T.,

RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,

RA Svirskaas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,

RA Wang Z.-Y., Wasserman D.A., Weinstein G.M., Weisenbach J.,

RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,

RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,

RA Zhang X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,

RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;

RT "The genome sequence of Drosophila melanogaster.";

RL Science 287:2185-2195(2000).

DR EMBL; AE003836; AAF59045.1.

DR FLYBase; FBgn027585; BCNA:GH05582.

SO SEQUENCE 985 AA; 112894 MW; 245F18B903DAC636 CRC64;

Query Match 37.5%; Score 6; DB 5; Length 985;  
Best Local Similarity 100.0%; Pred. No. 1.3e+02;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 VGYGRD 6  
Db 819 VGYGRD 824

## RESULT 29

ID Q8W078 PRELIMINARY; PRT; 1955 AA.

DT 01-MAR-2002 (TRENBLREL. 20, Created)

DT 01-MAR-2002 (TRENBLREL. 20, Last sequence update)

DT 01-MAR-2002 (TRENBLREL. 20, Last annotation update)

DE Helicase-like protein.

GN OSJNBA0091E23.15

OS Oryza sativa (Rice).

OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;

OC Erihartoideae; Oryzaceae; Oryza.

OX NCBI\_TaxID=4530;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=CV; NIPONBARE;

RA Sasaki T., Matsumoto T., Yamamoto K.;

RT "Oryza sativa nippobare (GA3) genomic DNA, chromosome 1, BAC

clone:OSJNBA0091E23.";

RL Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.

DR EMBL; AF003853; BAB84479.1;

SO SEQUENCE 1955 AA; 218758 MW; EE6A4CA5A8F42581 CRC64;

Query Match 37.5%; Score 6; DB 10; Length 1955;  
Best Local Similarity 100.0%; Pred. No. 2.4e+02;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 5 RDDNDR 10  
Db 608 RDDNDR 613

## RESULT 30

Q9W332

ID Q9W332 PRELIMINARY; PRT; 4215 AA.  
AC Q9W332;  
DT 01-MAY-2000 (TREMBLrel. 13, Created)  
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)  
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)  
DE CG2996 protein.  
GN CG2996 OR CG15318 OR CG18535.  
OS Drosophila melanogaster (Fruit fly).  
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
OC Ephydriidae; Drosophilidae; Drosophila.  
CX NCBI\_TaxID=7227;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=BERKELEY;  
RX MEDLINE=20196006; PubMed=10731132;  
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,  
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,  
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,  
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,  
RA Brandon R.C., Rogers Y.-H.C., Blazer R.G., Champagne M., Pfeiffer B.D.,  
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,  
RA Abail J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,  
RA Ballew R.M., Basu A., Baxendale J., Bayraktoglu L., Beasley E.M.,  
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,  
RA Borokova D., Botchan M.R., Bouck J., Brockstein P., Brotler P.,  
RA Butts K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,  
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,  
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,  
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,  
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,  
RA Fowler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,  
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,  
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,  
RA Hostin D., Houston K.A., Howland T.U., Wei M.-H., Ibegwam C.,  
RA Jaitai M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,  
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,  
RA Laoko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,  
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,  
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,  
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,  
RA Nelson D.R., Nelson K.A., Nixon K., Nusser D.R., Pauleb J.M.,  
RA Palazolo M., Peltman G.S., Pan S., Pollard J., Puri V., Reese M.G.,  
RA Reiner K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,  
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,  
RA Spier B., Spradling A.C., Stapleton M., Strong R., Sun E.,  
RA Svirskae R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,  
RA Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissbach J.,  
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,  
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,  
RA Zheng X.H., Zhong F.N., Zhou W., Zhou X., Zhu S., Zhu X., Smith H.O.,  
RA Glabbs R.A., Myers E.W., Rubin G.W., Venter J.C.;  
RT "The genome sequence of Drosophila melanogaster.";  
RL Science 287:2185-2195 (2000).  
CC -1 SIMILARITY: CONSTRAINTS 19 CUB DOMAINS.  
EMBL; AF003447; AAF46505.1; -  
HSSP; P01132; 1EGF  
DR FLYBase; FBgn0030138; CG2996.  
DR InterPro; IPR002123; Acyltransferase.  
DR InterPro; IPR000152; Asx\_hydroxyl.  
DR InterPro; IPR000859; CUB domain.  
DR InterPro; IPR002048; EF-hand.  
DR InterPro; IPR000561; EGF-like.  
DR InterPro; IPR000742; EGF\_2.  
DR InterPro; IPR001881; EGF\_Ca.  
DR Pfam; PF00431; CUB; 20.  
DR Pfam; PF00008; EGF; 6.  
DR SMART; SM00042; CUB; 19.  
DR SMART; SM00179; EGF\_CA; 2.  
DR SMART; SM00001; EGF\_like; 6.  
DR PROSITE; PS00100; ASX\_HYDROXYL; 2.  
DR PROSITE; PS01180; CUB; 29.

DR PROSITE; PS00022; EGF\_1; 3.  
DR PROSITE; PS01186; EGF\_2; 4.  
DR PROSITE; PS01187; EGF\_CA; 3.  
KW Calcium-binding; EGF-like domain; Glycoprotein; Hydroxylation; Repeat.  
SQ SEQUENCE 4215 AA; 470517 MW; 53613B810BA8D4F1 CRC64;  
Query Match 37.5%; Score 6; DB 5; Length 4215;  
Best Local Similarity 100.0%; Pred. No. 4.8e+02;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Qy 9 DRDPSR 14  
Db 3697 DRDPSR 3702  
RESULT 31  
ID Q9IWZ3 PRELIMINARY; PRT; 40 AA.  
AC Q9IWZ3;  
DT 01-OCT-2000 (TREMBLrel. 15, Created)  
DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)  
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)  
DE VP2 protein (Fragment).  
GN VP2.  
OS Human echovirus 7.  
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Picornaviridae;  
OC Enterovirus.  
CX NCBI\_TaxID=46018;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=1351NE98;  
RA Casas I., Palacios G.F., Trallero G., Cisterna D., Freire M.C.,  
RA Tenorio A.;  
RT "Molecular characterization of human enteroviruses in clinical samples  
by three different RT nested PCR assays and direct sequencing of  
RT amplified products.";  
RL submitted (APR-2000) to the EMBL/Genbank/DBJ databases.  
DR EMBL; AF252198; AAF71335.1; -  
DR InterPro; IPR001676; RVH.  
DR Pfam; PF00073; RVH; 1.  
FT NON\_TER 1  
FT NON\_TER 40  
SQ SEQUENCE 40 AA; 4457 MW; 8E898A8A779FEA6E CRC64;  
Query Match 31.2%; Score 5; DB 12; Length 40;  
Best Local Similarity 100.0%; Pred. No. 1e+02;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Qy 1 VGYGR 5  
Db 22 VGYGR 26  
RESULT 32  
ID Q9IYW8 PRELIMINARY; PRT; 40 AA.  
AC Q9IYW8;  
DT 01-OCT-2000 (TREMBLrel. 15, Created)  
DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)  
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)  
DE VP2 protein (Fragment).  
GN VP2.  
OS Human coxsackievirus A9.  
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Picornaviridae;  
OC Enterovirus.  
CX NCBI\_TaxID=12067;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=RI5797;  
RA Casas I., Palacios G.F., Trallero G., Cisterna D., Freire M.C.,  
RA Tenorio A.;  
RT "Molecular characterization of human enteroviruses in clinical samples  
by three different RT nested PCR assays and direct sequencing of

RT amplified products.";  
RL Submitted (APR-2000) to the EMBL/Genbank/DBJ databases.  
DR EMBL: AF2522203; AAF71340.1; -.  
DR InterPro: IPR001676; RNV.  
DR Pfam: PF00073; rhv; 1.  
FT NON\_TER 1  
FT NON\_TER 40  
SQ SEQUENCE 40 AA; 4445 MW; 60A48A9BEEF9BC3 CRC64;

Query Match 31.2%; Score 5; DB 12; Length 40;  
Best Local Similarity 100.0%; Pred. No. 1e+02;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 VGYGR 5  
|||||  
Db 22 VGYGR 26

RESULT 33  
Q9YUV4 PRELIMINARY; PRT; 50 AA.  
ID Q9YUV4  
AC Q9YUV4;  
DT 01-MAY-1999 (TREMBLrel. 10, Created)  
DT 01-MAY-1999 (TREMBLrel. 10, Last sequence update)  
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)  
DE VP2 (Fragment).  
GN Human echovirus 7.  
OS Viruses; ssRNA positive-strand viruses, no DNA stage; Picornaviridae;  
OC Enterovirus.  
OC NCBI\_TaxID=46018;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=WALLACE;  
RX MEDLINE=99094369; PubMed=9879760;  
RA Oberste M.S., Maher K., Pallansch M.A.;  
RT "Molecular phylogeny of all human enterovirus serotypes based on  
RT comparison of sequences at the 5' end of the region encoding VP2."  
RL Virus Res. 58:35-43(1998).  
DR EMBL: AF073767; AAD05295.1; -.  
DR InterPro: IPR001676; RNV.  
DR Pfam: PF00073; rhv; 1.  
FT NON\_TER 1  
FT NON\_TER 50  
SQ SEQUENCE 50 AA; 5548 MW; A38479F579C0E227 CRC64;

Query Match 31.2%; Score 5; DB 12; Length 50;  
Best Local Similarity 100.0%; Pred. No. 1.2e+02;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 VGYGR 5  
|||||  
Db 27 VGYGR 31

RESULT 34  
Q66826 PRELIMINARY; PRT; 51 AA.  
ID Q66826  
AC Q66826;  
DT 01-NOV-1996 (TREMBLrel. 01, Created)  
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)  
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)  
DE Capsid protein VP2 (Fragment).  
GN VP2.  
OS Human echovirus 5.  
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Picornaviridae;  
OC Enterovirus.  
OC NCBI\_TaxID=40280;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=NOYCE;  
RX MEDLINE=96203959; PubMed=8627260;  
RA Huttunen P., Santti J., Pulli T., Hyypia T.;  
RT "The major echovirus group is genetically coherent and related to

RT coxsackie B viruses.";  
RL J. Gen. Virol. 77:715-725(1996).  
DR EMBL: X89535; CA61713.1; -.  
DR InterPro: IPR001676; RNV.  
DR Pfam: PF00073; rhv; 1.  
FT NON\_TER 51  
FT NON\_TER 51  
SQ SEQUENCE 51 AA; 5563 MW; 9932C03CCF754D3E CRC64;

Query Match 31.2%; Score 5; DB 12; Length 51;  
Best Local Similarity 100.0%; Pred. No. 1.2e+02;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 VGYGR 5  
|||||  
Db 28 VGYGR 32

RESULT 35  
O8X2J9 PRELIMINARY; PRT; 63 AA.  
ID O8X2J9  
AC O8X2J9;  
DT 01-MAR-2002 (TREMBLrel. 20, Created)  
DT 01-MAR-2002 (TREMBLrel. 20, Last sequence update)  
DT 01-MAR-2002 (TREMBLrel. 20, Last annotation update)  
DE Hypothetical protein Ecs1375.  
GN Ecs1375.  
OS Escherichia coli O157:H7.  
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;  
OC Escherichia.  
OC NCBI\_TaxID=83334;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=O157:H7 / RIMD 0509952;  
RX MEDLINE=21156231; PubMed=11258796;  
RA Hayashi T., Makino K., Chitishi M., Kurokawa K., Ishii K., Yokoyama K.,  
RA Han C.-G., Ontsido E., Nakayama K., Murata T., Tanaka M., Tobe T.,  
RA Iida T., Takami H., Honda T., Sasakiwa C., Ogasawara N., Yasunaga T.,  
RA Kishara S., Shiba T., Hattori M., Shinagawa H.;  
RT "Complete genome sequence of enterohemorrhagic Escherichia coli  
RT O157:H7 and genomic comparison with a laboratory strain K-12.";  
RL DNA Res. 8:11-22(2001).  
DR EMBL: AP002554; BAB34798.1; -.  
DR Hypothetical protein.  
KW Hypothetical protein.  
SQ SEQUENCE 63 AA; 7551 MW; 5CDEA471D6BA8CFB CRC64;

Query Match 31.2%; Score 5; DB 16; Length 63;  
Best Local Similarity 100.0%; Pred. No. 1.5e+02;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 7 DNDRD 11  
|||||  
Db 12 DNDRD 16

RESULT 36  
O21995 PRELIMINARY; PRT; 75 AA.  
ID O21995  
AC O21995;  
DT 01-JAN-1998 (TREMBLrel. 05, Created)  
DT 01-JAN-1998 (TREMBLrel. 05, Last sequence update)  
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)  
DE ORF75 protein.  
OS Streptococcus thermophilus bacteriophage Sf121.  
OC Viruses; dsDNA viruses, no RNA stage; Caudovirales; Siphoviridae.  
OC NCBI\_TaxID=64186;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=96204576; PubMed=8623559;  
RA Bruttin A., Brusow H.;  
RT "Site-specific spontaneous deletions in three genome regions of a  
RT temperate Streptococcus thermophilus phage.";  
RL Virology 219:96-104(1996).  
RN [2]

RP SEQUENCE FROM N.A.  
RX MEDLINE=97344856; PubMed=9201223;  
RA Brutin A., Desiere F., Lucchini S., Foley S., Brusow H.;  
RT "Characterization of the lysogeny DNA module from the temperate  
ST streptococcus thermophilus bacteriophage phi Sfi21.";  
RU Virology 233:136-148(1997).  
RN [3]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=98160788; PubMed=9499809;  
RA Desiere F., Lucchini S., Brusow H.;  
RT "Evolution of Streptococcus thermophilus bacteriophage genomes by  
RT modular exchanges followed by point mutations and small deletions and  
RT insertions.";  
RU Virology 241:345-356(1998).  
RN [4]  
RP SEQUENCE FROM N.A.  
RA Desiere F., Lucchini S., Brutin A., Brusow H.;  
RT "Streptococcus thermophilus bacteriophage Sfi21 complete genome.";  
RU Submitted (DEC-1998) to the EMBL/GenBank/DBJ databases.  
DR EMBL; X95646; CA64936.1; -.  
DR EMBL; AF115103; AAD4099.1; -.  
DR InterPro; IPR001387; HTH\_3.  
DR Pfam; PF01381; HTH\_3; 1.  
DR SMART; SM00530; HTH\_XRE; 1.  
SQ SEQUENCE 75 AA; 8553 MW; B24E0861FAD0AAB4 CRC64;  
Query Match 31.2%; Score 5; DB 9; Length 75;  
Best Local Similarity 100.0%; Pred. No. 1.7e+02;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 2 GYGRD 6  
DB 56 GYGRD 60  
RESULT 37  
Q9KIL9 PRELIMINARY; PRT; 75 AA.  
ID Q9KIL9  
AC Q9KIL9;  
DT 01-OCT-2000 (TrEMBLrel. 15, Created)  
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)  
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)  
DE Hypothetical protein NMB0092.  
GN NMB0092.  
OS Neisseria meningitidis (serogroup B).  
OC Bacteria; Proteobacteria; beta subdivision; Neisseriaceae; Neisseria.  
OX NCBI\_TaxID=491;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX STRAIN=MC58 / SEROGROUP B;  
RC MEDLINE=2017575; PubMed=10710307;  
RA Tettein H., Saunders N.J., Heidelberg J., Jeffries A.C., Nelson K.E.,  
RA Eisen J.A., Ketchum K.A., Hood D.W., Peden J.F., Dodson R.U.,  
RA Nelson W.C., Gwin M.L., DeBoy R., Peterson J.D., Hickey E.K.,  
RA Haft D.H., Salzberg S.L., White O., Fleischmann R.D., Dougherty B.A.,  
RA Mason T., Ciecko A., Parksey D.S., Blatt E., Ciftone H., Clark E.B.,  
RA Cotton M.D., Uterback T.R., Khouri H., Qin H., Vamshanan J.,  
RA Gill J., Scarlato V., Massignani V., Pizza M., Grandi G., Sun L.,  
RA Smith H.O., Fraser C.M., Moxon E.R., Rappelli R., Venter J.C.;  
RT "Complete genome sequence of Neisseria meningitidis serogroup B strain  
RT MC58.";  
RU Science 287:1809-1815(2000).  
DR EMBL; AB002368; AAF40554.1; -.  
DR TIGR; NMB0092; -.  
KW Hypothetical protein; Complete proteome.  
SQ SEQUENCE 75 AA; 8504 MW; B60F7F15CC8ACFB8 CRC64;  
Query Match 31.2%; Score 5; DB 16; Length 75;  
Best Local Similarity 100.0%; Pred. No. 1.7e+02;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 12 PSRRN 16  
DB 56 PSRRN 16

DB 54 PSRRN 58  
RESULT 38  
Q9WY01 PRELIMINARY; PRT; 76 AA.  
ID Q9WY01  
AC Q9WY01;  
DT 01-NOV-1999 (TrEMBLrel. 12, Created)  
DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)  
DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)  
DE Hypothetical protein TM0154.  
GN TM0154.  
OS Thermotoga maritima.  
OC Bacteria; Thermotogae; Thermotogales; Thermotogaceae; Thermotoga.  
OX NCBI\_TaxID=2336;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX STRAIN=MSB8 / DSM 3109;  
RC MEDLINE=99287316; PubMed=10360571;  
RA Nelson K.E., Clayton R.A., Gill S.R., Gwin M.L., Dodson R.J.,  
RA Haft D.H., Hickey E.K., Peterson J.D., Nelson W.C., Ketchum K.A.,  
RA McDonald L., Uterback T.R., Malek J.A., Linher K.D., Garrett M.M.,  
RA Stewart A.M., Cotton M.D., Pratt M.S., Phillips C.A., Richardson D.,  
RA Heidelberg J., Sutton G.G., Fleischmann R.D., Eisen J.A., White O.,  
RA Salzberg S.L., Smith H.O., Venter J.C., Fraser C.M.;  
RT "Evidence for lateral gene transfer between Archaea and Bacteria from  
RT genome sequence of Thermotoga maritima.";  
RU Nature 398:323-329(1999).  
DR EMBL; AB001700; AAD35247.1; -.  
DR TIGR; TM0154; -.  
KW Hypothetical protein; Complete proteome.  
SQ SEQUENCE 76 AA; 9303 MW; 79F0C2D0D907082B CRC64;  
Query Match 31.2%; Score 5; DB 16; Length 76;  
Best Local Similarity 100.0%; Pred. No. 1.8e+02;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 12 PSRRN 16  
DB 27 PSRRN 31  
RESULT 39  
Q8XBR5 PRELIMINARY; PRT; 80 AA.  
ID Q8XBR5  
AC Q8XBR5;  
DT 01-MAR-2002 (TrEMBLrel. 20, Created)  
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)  
DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)  
DE Orf, hypothetical protein.  
GN Z3642 OR\_ECS3257.  
OS Escherichia coli O157:H7.  
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;  
OC Escherichia.  
OX NCBI\_TaxID=83334;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX STRAIN=O157:H7 / ATCC 700927;  
RC MEDLINE=0157:H7 / EDL933 / ATCC 700927;  
RA Pena N.T., Plunkett G., III, Burland V., Mau B., Glasner J.D.,  
RA Roe D.J., Mayhew G.F., Evans P.S., Gregor J., Kirkpatrick H.A.,  
RA Posfai G., Hackett J., Klink S., Boutin A., Shao Y., Miller L.,  
RA Grobbeck E.U., Davis N.W., Lim A., Dimantla E.T., Potanous K.,  
RA Apodaca J., Anantharaman T.S., Lin J., Yen G., Schwartz D.C.,  
RA Welch R.A., Blattner F.R.;  
RT "Genome sequence of enterohaemorrhagic Escherichia coli O157:H7.";  
RU Nature 409:529-533(2001).  
RN [2]  
RP SEQUENCE FROM N.A.  
RX STRAIN=O157:H7 / RIMD 0509952;  
RC MEDLINE=21156231; PubMed=11258796;  
RA Hayashi T., Makino K., Ohnishi M., Kurokawa K., Ishii K., Yokoyama K.,  
Han C.-G., Ohtsubo E., Nakayama K., Murata T., Tanaka M., Tobe T.,

RA Iida T., Takami H., Honda T., Sasakawa C., Ogasawara N., Yasunaga T.,  
 RA Kuhara S., Shiba T., Hattori M., Shinagawa H.,  
 RT "Complete genome sequence of enterohemorrhagic *Escherichia coli*  
 RT O157:H7 and genomic comparison with a laboratory strain K-12.";  
 RL DNA Res. 8:11-22(2001).  
 DR EMBL; AE005469; AAG57503.1; -;  
 DR EMBL; AF002561; BAB36680.1; -;  
 KM Complete proteome.  
 SQ SEQUENCE 80 AA; 8957 MW; 7835A2E8E3574F28 CRC64;

Query Match 31.2%; Score 5; DB 16; Length 80;  
 Best Local Similarity 100.0%; Pred. No. 1.8e+02;  
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VGYGR 5  
 |||||  
 Db 75 VGYGR 79

## RESULT 40

Q8VR70 PRELIMINARY; PRT; 86 AA.  
 AC Q8VR70;  
 DT 01-MAR-2002 (TrEMBLrel. 20. Created)  
 DT 01-MAR-2002 (TrEMBLrel. 20. Last sequence update)  
 DT 01-MAR-2002 (TrEMBLrel. 20. Last annotation update)  
 DE Insertion sequence protein.  
 OS *Escherichia coli*.  
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;  
 OC *Escherichia*.  
 OX NCB1\_TaxID=562;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=CFT073;  
 RX MEDLINE=21458820; PubMed=11574920;  
 RA Rasko D.A., Phillips J.A., Li X., Mobley H.L.T.;  
 RT "Identification of DNA sequences from a second pathogenicity island of  
 RT uropathogenic *Escherichia coli* CFT073: probes specific for  
 RT uropathogenic populations.";  
 RL J. Infect. Dis. 184:1041-1049(2001).  
 DR EMBL; AF447814; AAI67378.1; -;  
 SQ SEQUENCE 86 AA; 9826 MW; 891469A73C1B7A77 CRC64;

Query Match 31.2%; Score 5; DB 2; Length 86;  
 Best Local Similarity 100.0%; Pred. No. 2e+02;  
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VGYGR 5  
 |||||  
 Db 34 VGYGR 38

Search completed: April 9, 2003, 13:15:35  
 Job time : 39.9016 secs

GenCore version 5.1.4 p5.4578  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: April 8, 2003, 19:24:10 ; Search time 13.377 Seconds  
(without alignments)  
1049.161 Million cell updates/sec

Title: US-10-020-441-3  
Perfect score: 90  
Sequence: 1 VGYGRDNDNRDPSRKN 16

Scoring table:  
BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 6.0 , Ygapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 593429 seqs, 438583890 residues

Total number of hits satisfying chosen parameters: 1186858

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Command line parameters:  
-MODEL=frame+ p2n.model -DBY=xjh  
-Q=/cg21/USPTO.spool/US10020441/rnuc 02042003 093014 21095/app\_query.fasta\_1.462  
-DB=Published.Applications.NA -QFW=faetap -SUFIX=trpb -MINMATCH=0.1  
-LOOPCU=0 -LOOPEXT=0 -UNITS=bases -START=1 -MATRIX=blonsum62  
-TRANS=human40.cdi -LIST=45 -DOCALIGN=200 -THR SCORE=pct -THR MAX=100  
-THR MIN=0 -ALIGN=40 -MODE=LOCAL -OUTFMT=pct -NORM=ext -HEAPSIZE=500 -MINLEN=0  
-MAXLEN=200000000 -USER=US10020441 @CGN 1.1 89 @rnuc 02042003 093014 21095  
-NCPU=6 -ICPU=3 -NO\_XLIPYX -NO\_MMAP -LARGEQUERY -NEG\_SCORES=0 -WAIT -LONGLOG  
-DEV\_TIMEOUT=120 -MAIN\_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6  
-FGAPEXT=7 -YGAPEXT=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

- Published Applications NA:\*
- 1: /cg21\_6/ptodata/1/pubpna/US07\_PUBCOMB.seq:\*
  - 2: /cg21\_6/ptodata/1/pubpna/PCT\_NEW\_PUB.seq:\*
  - 3: /cg21\_6/ptodata/1/pubpna/US06\_NEW\_PUB.seq:\*
  - 4: /cg21\_6/ptodata/1/pubpna/US06\_PUBCOMB.seq:\*
  - 5: /cg21\_6/ptodata/1/pubpna/US07\_NEW\_PUB.seq:\*
  - 6: /cg21\_6/ptodata/1/pubpna/PCTUS\_PUBCOMB.seq:\*
  - 7: /cg21\_6/ptodata/1/pubpna/US08\_NEW\_PUB.seq:\*
  - 8: /cg21\_6/ptodata/1/pubpna/US08\_PUBCOMB.seq:\*
  - 9: /cg21\_6/ptodata/1/pubpna/US09\_NEW\_PUB.seq:\*
  - 10: /cg21\_6/ptodata/1/pubpna/US09\_PUBCOMB.seq:\*
  - 11: /cg21\_6/ptodata/1/pubpna/US10\_NEW\_PUB.seq:\*
  - 12: /cg21\_6/ptodata/1/pubpna/US10\_PUBCOMB.seq:\*
  - 13: /cg21\_6/ptodata/1/pubpna/US60\_NEW\_PUB.seq:\*
  - 14: /cg21\_6/ptodata/1/pubpna/US60\_PUBCOMB.seq:\*

\*Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length DB	ID	Description
1	90	100.0	1477	9	US-10-020-441-1
2	52	57.8	2112	10	US-09-529-063-77
3	52	57.8	2690	9	US-09-822-846-132
4	52	57.8	2880	10	US-09-740-516-1

C 5	52	57.8	3492	10	US-09-529-063-78	Sequence 78, Appl
C 6	46	51.1	138	10	US-09-864-761-27520	Sequence 21520, A
C 7	46	51.1	410	10	US-09-864-761-10881	Sequence 10881, A
C 8	46	51.1	453	10	US-09-960-352-4774	Sequence 4774, Ap
C 9	45.5	50.6	381	10	US-09-960-352-9767	Sequence 9767, Ap
C 10	44	48.9	1029	10	US-09-770-445-193	Sequence 193, App
C 11	44	48.9	1842	9	US-10-151-557-1	Sequence 1, Appl1
C 12	44	48.9	3614	10	US-09-880-107-1753	Sequence 1753, Ap
C 13	43	47.8	358	10	US-09-960-352-14356	Sequence 14356, A
C 14	43	47.8	398	10	US-09-960-352-6155	Sequence 6155, Ap
C 15	43	47.8	759	9	US-09-738-626-37	Sequence 37, Appl
C 16	43	47.8	1149	10	US-09-547-267-8	Sequence 8, Appl1
C 17	43	47.8	1699	12	US-10-002-600-62	Sequence 62, Appl
C 18	43	47.8	1700	9	US-10-098-841-329	Sequence 133, App
C 19	43	47.8	1737	10	US-09-764-864-133	Sequence 1, Appl1
C 20	43	47.8	8625	9	US-09-920-923-1	Sequence 2259, Ap
C 21	43	47.8	9745	9	US-10-091-504-2259	Sequence 2259, Ap
C 22	43	47.8	9745	10	US-09-764-869-2259	Sequence 27, Appl
C 23	43	47.8	11233	9	US-09-920-923-27	Sequence 2258, Ap
C 24	43	47.8	12149	9	US-10-091-504-2258	Sequence 2258, Ap
C 25	43	47.8	12149	10	US-09-764-869-2258	Sequence 1, Appl1
C 26	43	47.8	3309400	9	US-09-738-626-1	Sequence 7832, Ap
C 27	42	46.7	261	10	US-09-878-574-7832	Sequence 59, Appl
C 28	42	46.7	918	9	US-10-029-180-59	Sequence 575, App
C 29	42	46.7	1742	10	US-09-764-864-575	Sequence 29, Appl
C 30	42	46.7	1851	9	US-10-047-260-35	Sequence 35, Appl
C 31	42	46.7	1884	9	US-10-047-260-35	Sequence 1, Appl1
C 32	42	46.7	2372	10	US-09-752-983-1	Sequence 1, Appl1
C 33	42	46.7	2372	10	US-09-956-425-7	Sequence 1, Appl1
C 34	42	46.7	2372	10	US-09-851-771A-1	Sequence 493, App
C 35	42	46.7	4108	9	US-09-925-301-493	Sequence 165, App
C 36	41	45.6	438	9	US-10-101-464A-165	Sequence 11028, A
C 37	41	45.6	462	10	US-09-864-761-11028	Sequence 224, App
C 38	41	45.6	861	9	US-09-764-868-234	Sequence 699, App
C 39	41	45.6	1209	10	US-09-974-300-699	Sequence 24, Appl
C 40	41	45.6	3027	9	US-10-125-540-24	Sequence 24, Appl
C 41	41	45.6	3027	10	US-09-764-870-24	Sequence 3, Appl1
C 42	41	45.6	19740	9	US-10-142-356-3	Sequence 3, Appl1
C 43	41	45.6	143306	10	US-09-729-920-3	Sequence 6813, Ap
C 44	40.5	45.0	732	10	US-09-974-300-6813	Sequence 6646, Ap
C 45	40.5	45.0	867	10	US-09-974-300-6646	

## ALIGNMENTS

RESULT 1  
US-10-020-441-1  
; Sequence 1, Application us/10020441  
; Publication No. US2002018224A1  
GENERAL INFORMATION:  
APPLICANT: Doenhoff, Michael  
APPLICANT: Sayers, Jon  
TITLE OF INVENTION: SCHISTOSOMIASIS VACCINE  
FILE REFERENCE: Doenhoff  
CURRENT APPLICATION NUMBER: US/10/020,441  
PRIOR FILING DATE: 2001-12-18  
PRIOR APPLICATION NUMBER: US/09/413,810  
PRIOR FILING DATE: RPTOR FILING DATE: 1999-10-07  
NUMBER OF SEQ ID NOS: 3  
SOFTWARE: PatentIn Ver. 2.0  
SEQ ID NO 1  
LENGTH: 1477  
TYPE: DNA  
ORGANISM: Schistosoma mansoni  
US-10-020-441-1

Alignment Scores:  
Pred. No.: 2.3e-07  
Score: 90.00  
Percent Similarity: 100.00%  
Best Local Similarity: 100.00%  
Query Match: 9  
DB: 9  
Length: 1477  
Matches: 16  
Conservative: 0  
Mismatch: 0  
Indels: 0  
Gaps: 0





```
FILE REFERENCE: Q58769
CURRENT APPLICATION NUMBER: US/09/529,063
CURRENT FILING DATE: 2000-04-07
PRIOR APPLICATION NUMBER: PCT/JP98/04514
PRIOR FILING DATE: 1998-10-06
PRIOR APPLICATION NUMBER: JP 9-274674
PRIOR FILING DATE: 1997-10-07
NUMBER OF SEQ ID NOS: 117
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 78
LENGTH: 3492
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: sig_deptide
LOCATION: (6)..(53)
NAME/KEY: mac_deptide
LOCATION: (54)..(2117)
NAME/KEY: CDS
LOCATION: (6)..(2117)
US-09-529-063-78

Alignment Scores:
Pred. No.: 3.83 Length: 3492
Score: 52.00 Matches: 11
Percent Similarity: 85.71% Conservative: 1
Best Local Similarity: 78.57% Mismatches: 1
Query Match: 57.78% Indels: 1
DB: 10 Gaps: 0

US-10-020-441-3 (1-16) x US-09-529-063-78 (1-3492)
Qy 1 ValGIYrGIYrGIYrGAspAspAspAsp-ArgAspProSer 13
Db 1380 ATGGGATrGGTACTGACGACATGATCATGAGACCCCAAGT 1341

RESULT 6
US-09-864-761-27520/c
Sequence 27520, Application US/09864761
Patent No. US20020048763A1
GENERAL INFORMATION:
APPLICANT: Penn, Sharon G.
APPLICANT: Rank, David R.
APPLICANT: Hanzel, David K.
APPLICANT: Chen, Wensheng
TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
FILE REFERENCE: Aeomica-X-1
CURRENT APPLICATION NUMBER: US/09/864,761
CURRENT FILING DATE: 2001-05-23
PRIOR APPLICATION NUMBER: US 60/180,312
PRIOR FILING DATE: 2000-02-04
PRIOR APPLICATION NUMBER: US 60/207,456
PRIOR FILING DATE: 2000-05-26
PRIOR APPLICATION NUMBER: US 09/632,366
PRIOR FILING DATE: 2000-08-03
PRIOR APPLICATION NUMBER: GB 24263,6
PRIOR FILING DATE: 2000-10-04
PRIOR APPLICATION NUMBER: US 60/236,359
PRIOR FILING DATE: 2000-09-27
PRIOR APPLICATION NUMBER: PCT/US01/00666
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00667
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00664
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00669
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00665
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00668
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00663
```

```
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00662
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00661
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00670
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: US 60/234,687
PRIOR FILING DATE: 2000-09-21
PRIOR APPLICATION NUMBER: US 09/608,408
PRIOR FILING DATE: 2000-06-30
PRIOR APPLICATION NUMBER: US 09/774,203
PRIOR FILING DATE: 2001-01-29
NUMBER OF SEQ ID NOS: 49117
SOFTWARE: Annonax Sequence Listing Engine vers. 1.1
SEQ ID NO 27520
LENGTH: 138
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
OTHER INFORMATION: MAP TO AC003683.1
OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 0.7
OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 0.73
OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 0.59
OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 0.68
OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 0.87
OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 0.6
OTHER INFORMATION: NT HIT: L35321.2, EVALUE 2.00e-02
OTHER INFORMATION: EST_HUMAN HIT: AW856789.1, EVALUE 1.20e-01
US-09-864-761-27520

Alignment Scores:
Pred. No.: 1.67 Length: 138
Score: 46.00 Matches: 9
Percent Similarity: 66.67% Conservative: 1
Best Local Similarity: 60.00% Mismatches: 5
Query Match: 51.11% Indels: 0
DB: 10 Gaps: 0

US-10-020-441-3 (1-16) x US-09-864-761-27520 (1-138)
Qy 2 G1YrGIYrGIYrGAspAspAspAsp-ArgAspProSerArgLysAsn 16
Db 123 GGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTCTAATAAAGC 79

RESULT 7
US-09-864-761-10881/c
Sequence 10881, Application US/09864761
Patent No. US20020048763A1
GENERAL INFORMATION:
APPLICANT: Penn, Sharon G.
APPLICANT: Rank, David R.
APPLICANT: Hanzel, David K.
APPLICANT: Chen, Wensheng
TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
FILE REFERENCE: Aeomica-X-1
CURRENT APPLICATION NUMBER: US/09/864,761
CURRENT FILING DATE: 2001-05-23
PRIOR APPLICATION NUMBER: US 60/180,312
PRIOR FILING DATE: 2000-02-04
PRIOR APPLICATION NUMBER: US 60/207,456
PRIOR FILING DATE: 2000-05-26
PRIOR APPLICATION NUMBER: US 09/632,366
PRIOR FILING DATE: 2000-08-03
PRIOR APPLICATION NUMBER: GB 24263,6
PRIOR FILING DATE: 2000-10-04
PRIOR APPLICATION NUMBER: US 60/236,359
PRIOR FILING DATE: 2000-09-27
PRIOR APPLICATION NUMBER: PCT/US01/00666
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00667
PRIOR FILING DATE: 2001-01-30
```

```

; SEQ ID NO 4774
;
; LENGTH: 453
;
; TYPE: DNA
;
; ORGANISM: Bos taurus
;
; FEATURE:
;
; NAME/KEY: unsure

```

APPLICANT: Garcia, Carlos A.

USE OF INVENTION: NUCLEIC ACID AND OTHER MOLECULES ASSOCIATED WITH LACTATION AND

```

; APPLICANT: Krickler, Maja
; APPLICANT: Slader, Ted
; APPLICANT: Davis, Keith R.
; APPLICANT: Allen, Keith
; APPLICANT: Hoffman, Neil
; APPLICANT: Huidan, Patrick
; TITLE OF INVENTION: Expressed Sequences of Arabidopsis
; FILE REFERENCE: 2023US (PARA-012PRV)
; CURRENT APPLICATION NUMBER: US/09/770,445
; PRIOR FILING DATE: 2001-01-26
; PRIOR APPLICATION NUMBER: US 60/178,472
; NUMBER OF SEQ ID NOS: 999
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 193
; LENGTH: 1029
; TYPE: DNA
; ORGANISM: Arabidopsis thaliana
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1)..(1029)
; OTHER INFORMATION: n = A,T,C or G
US-09-770-445-193

Alignment Scores:
Pred. No.: 30.1 Length: 1029
Score: 44.00 Matches: 7
Percent Similarity: 66.67% Conservative: 3
Best Local Similarity: 46.67% Mismatches: 5
Query Match: 48.89% Indels: 0
DB: 10 Gaps: 0

US-10-020-441-3 (1-16) x US-09-770-445-193 (1-1029)

Cy 2 G|Y|Y|G|Y|A|G|A|P|A|S|P|A|S|P|A|S|P|P|S|E|A|T|G|Y|S|A|N 16
|||::||| |||::||| |||::||| |||::|||
Db 367 GGAATTGGGAATTGTGACCCGAGAGAGATCCGAAAGGTACAC 323

RESULT 11
US-10-151-557-1
; Sequence 1, Application US/10151557
; Publication No. US20020183506A1
; GENERAL INFORMATION:
; APPLICANT: No. US20020183506A1ozymes A/S
; APPLICANT: Danielsen, Steffen
; APPLICANT: Schneider, Paile
; TITLE OF INVENTION: Nucleic acids encoding polypeptides having
; FILE REFERENCE: 10173.200-US
; CURRENT APPLICATION NUMBER: US/10/151,557
; CURRENT FILING DATE: 2002-05-18
; PRIOR APPLICATION NUMBER: US/09/832,614A
; PRIOR FILING DATE: 2001-04-11
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 1
; LENGTH: 1842
; TYPE: DNA
; ORGANISM: Geniculisporium sp.
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(1842)
; OTHER INFORMATION:
US-10-151-557-1

Alignment Scores:
Pred. No.: 54.6 Length: 1842
Score: 44.00 Matches: 8
Percent Similarity: 90.00% Conservative: 1
Best Local Similarity: 80.00% Mismatches: 1
Query Match: 48.89% Indels: 0
DB: 9 Gaps: 0
```

```

US-10-020-441-3 (1-16) x US-10-151-557-1 (1-1842)

Cy 5 A|G|A|P|A|S|P|A|S|P|A|S|P|P|S|E|A|T|G|Y|S|A|N 14
|||::||| |||::||| |||::||| |||::|||
Db 1096 CGTGACGACAACTTCGCGACCAATCGC 1125

RESULT 12
US-09-880-107-1753
; Sequence 1753, Application US/09880107
; Patent No. US20020142981A1
; GENERAL INFORMATION:
; APPLICANT: Horne, Darci T.
; APPLICANT: Vockley, Joseph G.
; APPLICANT: Scherf, Uwe
; APPLICANT: Gene Logic, Inc.
; TITLE OF INVENTION: Gene Expression Profiles in Liver Cancer
; FILE REFERENCE: 44921-5028-WO
; CURRENT APPLICATION NUMBER: US/09/880,107
; CURRENT FILING DATE: 2001-06-14
; PRIOR APPLICATION NUMBER: US 60/211,379
; PRIOR FILING DATE: 2000-06-14
; PRIOR APPLICATION NUMBER: US 60/237,054
; PRIOR FILING DATE: 2000-10-02
; NUMBER OF SEQ ID NOS: 3950
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1753
; LENGTH: 3614
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: Genbank Accession No. US20020142981A1 D86956
US-09-880-107-1753

Alignment Scores:
Pred. No.: 109 Length: 3614
Score: 44.00 Matches: 8
Percent Similarity: 71.43% Conservative: 2
Best Local Similarity: 57.14% Mismatches: 4
Query Match: 48.89% Indels: 0
DB: 10 Gaps: 0

US-10-020-441-3 (1-16) x US-09-880-107-1753 (1-3614)

Cy 2 G|Y|Y|G|Y|A|G|A|P|A|S|P|A|S|P|A|S|P|P|S|E|A|T|G|Y|S|A|N 15
|||::||| |||::||| |||::||| |||::|||
Db 289 GCGAGGGGGCGGAGACGACGAGACCGAGCGGAGG 330

RESULT 13
US-09-960-352-14356/C
; Sequence 14356, Application US/09960352
; Patent No. US20020137139A1
; GENERAL INFORMATION:
; APPLICANT: Warren, Wesley C.
; APPLICANT: Tao, Nengbing
; APPLICANT: Byatt, John C.
; APPLICANT: Mathiasagan, Nagappan
; TITLE OF INVENTION: NUCLEIC ACID AND OTHER MOLECULES ASSOCIATED WITH LACTATION AND
; FILE REFERENCE: 16511.006/37-21(10298)C
; CURRENT APPLICATION NUMBER: US/09/960,352
; CURRENT FILING DATE: 2001-09-24
; NUMBER OF SEQ ID NOS: 15112
; SEQ ID NO 14356
; LENGTH: 358
; TYPE: DNA
; ORGANISM: Bos taurus
; OTHER INFORMATION: Clone ID: 61-LIB34-028-Q1-E1-H2
US-09-960-352-14356

Alignment Scores:
Pred. No.: 15.4 Length: 358
Score: 43.00 Matches: 7
```

Percent Similarity: 72.73% Conservative: 1  
Best Local Similarity: 63.64% Mismatches: 3  
Query Match: 47.78% Indels: 0  
DB: 10 Gaps: 0

US-10-020-441-3 (1-16) x US-09-960-352-14356 (1-358)

Qy 2 G1YTYrG1YArgAspAsnAspArgAspPro 12  
Db 42 GCCTTCTGGAGAGATGATGAGACAGAGACCA 10

RESULT 14  
US-09-960-352-6155/c

; Sequence 6155, Application US/09960352  
; Patent No. US20020137139A1  
; GENERAL INFORMATION:  
; APPLICANT: Warren, Wesley C.  
; APPLICANT: Tao, Nengping  
; APPLICANT: Byatt, John C.  
; APPLICANT: Mathialagan, Nagappan  
; TITLE OF INVENTION: NUCLEIC ACID AND OTHER MOLECULES ASSOCIATED WITH LACTATION AND  
; TITLE OF INVENTION: MUSCLE AND FAT DEPOSITION  
; FILE REFERENCE: 16511.006/37-21(10298)C  
; CURRENT APPLICATION NUMBER: US/09/960,352  
; CURRENT FILING DATE: 2001-09-24  
; NUMBER OF SEQ ID NOS: 15112  
; SEQ ID NO 6155  
; LENGTH: 398  
; TYPE: DNA  
; ORGANISM: Bos taurus  
; OTHER INFORMATION: Clone ID: 26-LIB34-079-Q1-E1-G9  
US-09-960-352-6155

Alignment Scores:

Pred. No.: 17.2 Length: 398  
Score: 43.00 Matches: 7  
Percent Similarity: 72.73% Conservative: 1  
Best Local Similarity: 63.64% Mismatches: 3  
Query Match: 47.78% Indels: 0  
DB: 10 Gaps: 0

US-10-020-441-3 (1-16) x US-09-960-352-6155 (1-398)

Qy 2 G1YTYrG1YArgAspAsnAspArgAspPro 12

Db 134 GCCTTCTGGAGAGATGATGAGACAGAGACCA 102

RESULT 15  
US-09-738-626-37/c

; Sequence 37, Application US/09738626  
; Publication No. US20020197605A1  
; GENERAL INFORMATION:  
; APPLICANT: NAKAGAWA, SATOSHI  
; APPLICANT: MIZOGUCHI, HIROSHI  
; APPLICANT: ANDO, SEIKO  
; APPLICANT: HAYASHI, MIKIRO  
; APPLICANT: OCHIAI, KEIKO  
; APPLICANT: YOKOI, HARUHIKO  
; APPLICANT: TATEISHI, NAOKO  
; APPLICANT: SENOH, AKIHIRO  
; APPLICANT: IKEDA, MASATO  
; APPLICANT: OZAKI, AKIO  
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES  
; FILE REFERENCE: 248-125  
; CURRENT APPLICATION NUMBER: US/09/738,626  
; CURRENT FILING DATE: 2000-12-18  
; PRIOR APPLICATION NUMBER: JP 99/377484  
; PRIOR FILING DATE: 1999-12-16  
; PRIOR APPLICATION NUMBER: JP 00/159162  
; PRIOR FILING DATE: 2000-04-07  
; PRIOR APPLICATION NUMBER: JP 00/280988  
; PRIOR FILING DATE: 2000-08-03  
; NUMBER OF SEQ ID NOS: 7059

; SOFTWARE: PatentIn ver. 3.0  
; SEQ ID NO 37  
; LENGTH: 759  
; TYPE: DNA  
; ORGANISM: Corynebacterium glutamicum  
US-09-738-626-37

Alignment Scores:

Pred. No.: 33.3 Length: 759  
Score: 43.00 Matches: 8  
Percent Similarity: 75.00% Conservative: 1  
Best Local Similarity: 66.67% Mismatches: 3  
Query Match: 47.78% Indels: 0  
DB: 9 Gaps: 0

US-10-020-441-3 (1-16) x US-09-738-626-37 (1-759)

Qy 1 ValG1YTYrG1YArgAspAsnAspArgAspPro 12  
Db 609 GTGGGGGTTGTGCGTGAAGAAATGACCGCGATCCC 574

RESULT 16  
US-09-547-267-8/c

; Sequence 8, Application US/09547267  
; Patent No. US20020147371A1  
; GENERAL INFORMATION:  
; APPLICANT: Hohmann, Hans-Peter  
; APPLICANT: Pasamontes, Luis  
; APPLICANT: Tessier, Michel  
; APPLICANT: van Loon, Adolphus  
; TITLE OF INVENTION: FERMENTATIVE CAROTENOID PRODUCTION  
; NUMBER OF SEQUENCES: 47  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Hoffmann-La Roche Inc.  
; STREET: 340 Kingsland Street  
; CITY: Nutley  
; STATE: NJ  
; COUNTRY: USA  
; ZIP: 07110  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/547,267  
; FILING DATE:  
; CLASSIFICATION:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 08/660,645  
; FILING DATE:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Pokras, Bruce A.  
; REGISTRATION NUMBER: 32,748  
; REFERENCE/DOCKET NUMBER: RAN 6002/170  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (201) 235-5801  
; TELEFAX: (201) 235-2363  
; INFORMATION FOR SEQ ID NO: 8:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1149 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: double  
; TOPOLOGY: linear  
; MOLECULE TYPE: DNA (genomic)  
US-09-547-267-8

Alignment Scores:  
Pred. No.: 50.9 Length: 1149  
Score: 43.00 Matches: 9  
Percent Similarity: 69.23% Conservative: 0  
Best Local Similarity: 69.23% Mismatches: 4  
Query Match: 47.78% Indels: 0

```
DB: 10 Gaps: 0
US-10-020-441-3 (1-16) x US-09-547-267-8 (1-1149)
Qy 2 G1YrGAspAspAspAspAspAspAspProSerArg14
Db 925 GGACGGGTGGGATCCGCGGATGATGCGCCAGCCGC 887

RESULT 17
US-10-002-600-62
; Sequence 62, Application US/10002600
; Patent No. US20020137077A1
; GENERAL INFORMATION:
; APPLICANT: Hopkins, Christopher M.
; APPLICANT: Peterson, David P.
; APPLICANT: Cocks, Benjamin G.
; APPLICANT: Hawkins, Phillip R.
; TITLE OF INVENTION: GENES REGULATED IN ACTIVATED T CELLS
; FILE REFERENCE: PA-0042 US
; CURRENT APPLICATION NUMBER: US/10/002,600
; PRIOR FILING DATE: 2001-10-25
; PRIOR APPLICATION NUMBER: 60/243,521
; PRIOR FILING DATE: 2000-10-25
; NUMBER OF SEQ ID NOS: 116
; SOFTWARE: PERL Program
; SEQ ID NO 62
; LENGTH: 1699
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; OTHER INFORMATION: Template ID: 238849.1
US-10-002-600-62

Alignment Scores:
Pred. No.: 76.1 Length: 1699
Score: 43.00 Matches: 8
Percent Similarity: 75.00% Conservative: 1
Best Local Similarity: 66.67% Mismatches: 3
Query Match: 47.78% Indels: 0
Gaps: 0
DB: 12

US-10-020-441-3 (1-16) x US-10-002-600-62 (1-1699)
Qy 4 G1YrGAspAspAspAspAspAspProSerArg15
Db 619 GGCGGTGGACCCGACCGAGAGTGAGAAA 654

RESULT 18
US-10-098-841-329
; Sequence 329, Application US/10098841
; Publication No. US20020197679A1
; GENERAL INFORMATION:
; APPLICANT: Tang, Y. Tom
; APPLICANT: Liu, Chenghua
; APPLICANT: Asundi, Vinod
; APPLICANT: Xu, Chongjun
; APPLICANT: Zhou, Ping
; APPLICANT: Ma, Yungqing
; APPLICANT: Wang, Jian-Rui
; APPLICANT: Zhao, Qing A.
; APPLICANT: Ren, Feiyan
; APPLICANT: Chen, Rui-hong
; APPLICANT: Wang, Dunui
; APPLICANT: Wang, Zhiwei
; APPLICANT: Wehrman, Tom
; APPLICANT: Zhang, Jie
; APPLICANT: Qian, Xiaohong B.
; APPLICANT: Drmanac, Radoje T.
; TITLE OF INVENTION: No. US20020197679A1el Nucleic Acids and
; FILE REFERENCE: 784C1P2
; CURRENT APPLICATION NUMBER: US/10/098, 841
```

```
; CURRENT FILING DATE: 2002-03-13
; PRIOR APPLICATION NUMBER: 09/598,042
; PRIOR FILING DATE: 2000-06-20
; PRIOR APPLICATION NUMBER: 09/552,317
; PRIOR FILING DATE: 2000-04-25
; PRIOR APPLICATION NUMBER: 09/488,725
; PRIOR FILING DATE: 2000-01-21
; NUMBER OF SEQ ID NOS: 331
; SOFTWARE: PL_FL_genes Version 1.0
; SEQ ID NO 329
; LENGTH: 1700
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (120)..(1553)
US-10-098-841-329

Alignment Scores:
Pred. No.: 76.1 Length: 1700
Score: 43.00 Matches: 8
Percent Similarity: 75.00% Conservative: 1
Best Local Similarity: 66.67% Mismatches: 3
Query Match: 47.78% Indels: 0
Gaps: 0
DB: 9

US-10-020-441-3 (1-16) x US-10-098-841-329 (1-1700)
Qy 4 G1YrGAspAspAspAspAspAspProSerArg15
Db 612 GGCGGTGGACCCGACCGAGAGTGAGAAA 647

RESULT 19
US-09-764-864-133
; Sequence 133, Application US/09764864
; Patent No. US20020132753A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PT223
; CURRENT APPLICATION NUMBER: US/09/764,864
; PRIOR FILING DATE: 2001-01-17
; PRIOR application data removed - consult PAM or file wrapper
; NUMBER OF SEQ ID NOS: 1792
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 133
; LENGTH: 1737
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-764-864-133

Alignment Scores:
Pred. No.: 77.8 Length: 1737
Score: 43.00 Matches: 8
Percent Similarity: 75.00% Conservative: 1
Best Local Similarity: 66.67% Mismatches: 3
Query Match: 47.78% Indels: 0
Gaps: 0
DB: 10

US-10-020-441-3 (1-16) x US-09-764-864-133 (1-1737)
Qy 4 G1YrGAspAspAspAspAspAspProSerArg15
Db 625 GGCGGTGGACCCGACCGAGAGTGAGAAA 660

RESULT 20
US-09-920-923-1
; Sequence 1, Application US/09920923
; Publication No. US20030022273A1
; GENERAL INFORMATION:
; APPLICANT: Pasamontes, Luis
; APPLICANT: Teygankov, Yuri
; TITLE OF INVENTION: Improved Fermentative Carotenoid Production
```

```
/ FILE REFERENCE: Improved Fermentative Carotenoid
/ CURRENT APPLICATION NUMBER: US/09/920,923
/ SEQUENCE FILING DATE: 2001-08-02
/ PRIOR APPLICATION NUMBER: 08/980,832
/ PRIOR FILING DATE: 1997-12-01
/ NUMBER OF SEQ ID NOS: 66
/ SOFTWARE: PatentIn Ver. 2.1
/ SEQ ID NO 1
/ LENGTH: 8625
/ TYPE: DNA
/ ORGANISM: Flavobacterium sp. R1534
/ FEATURE:
/ NAME/KEY: unsure
/ LOCATION: (8348)..(8349)
/ NAME/KEY: unsure
/ LOCATION: (8539)..(8540)
/ NAME/KEY: unsure
/ LOCATION: (8581)
/ NAME/KEY: unsure
/ LOCATION: (8590)
/ NAME/KEY: unsure
/ LOCATION: (8592)
/ NAME/KEY: unsure
/ LOCATION: (8602)..(8604)
/ LOCATION: (8602)..(8604)
US-09-920-923-1
```

## Alignment Scores:

```
Pred. No.: 403 Length: 8625
Score: 43.00 Matches: 9
Percent Similarity: 69.23% Conservative: 0
Best Local Similarity: 69.23% Mismatches: 4
Query Match: 47.78% Indels: 0
DB: 9 Gaps: 0
```

US-10-020-441-3 (1-16) x US-09-920-923-1 (1-8625)

Qy 2 G1YrG1YrGAspAspAspAspArgArg 14

Db 6018 GGAAGCGGTGGGATCGCGGCGATCGATGCCACGCCG 6056

## RESULT 21

```
US-10-091-504-2259/c
/ Sequence 2259, Application US/10091504
/ Publication No. US2003005908A1
/ GENERAL INFORMATION:
/ APPLICANT: Rosen et al.
/ TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
/ FILE REFERENCE: PC007C1
/ CURRENT APPLICATION NUMBER: US/10/091,504
/ CURRENT FILING DATE: 2002-03-07
/ NUMBER OF SEQ ID NOS: 2442
/ Prior Application removed - See File Wrapper or Palm
/ SOFTWARE: PatentIn Ver. 2.0
/ SEQ ID NO 2259
/ LENGTH: 9745
/ TYPE: DNA
/ ORGANISM: Homo sapiens
US-10-091-504-2259
```

## Alignment Scores:

```
Pred. No.: 457 Length: 9745
Score: 43.00 Matches: 7
Percent Similarity: 75.00% Conservative: 2
Best Local Similarity: 58.33% Mismatches: 3
Query Match: 47.78% Indels: 0
DB: 9 Gaps: 0
```

US-10-020-441-3 (1-16) x US-10-091-504-2259 (1-9745)

Qy 4 G1YrGAspAspAspAspArgArgProSerArgLys 15

Db 6855 GGCAGTGCATGGGACGACGAGCCAGAGAGAA 6820

## RESULT 22

```
US-09-764-869-2259/c
/ Sequence 2259, Application US/09764869
/ Patent No. US20020061521A1
/ GENERAL INFORMATION:
/ APPLICANT: Rosen et al.
/ TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
/ FILE REFERENCE: PC007
/ CURRENT APPLICATION NUMBER: US/09/764,869
/ SEQUENCE FILING DATE: 2001-01-17
/ Prior application data removed - refer to PALM or file wrapper
/ NUMBER OF SEQ ID NOS: 2442
/ SOFTWARE: PatentIn Ver. 2.0
/ SEQ ID NO 2259
/ LENGTH: 9745
/ TYPE: DNA
/ ORGANISM: Homo sapiens
US-09-764-869-2259
```

## Alignment Scores:

```
Pred. No.: 457 Length: 9745
Score: 43.00 Matches: 7
Percent Similarity: 75.00% Conservative: 2
Best Local Similarity: 58.33% Mismatches: 3
Query Match: 47.78% Indels: 0
DB: 10 Gaps: 0
```

US-10-020-441-3 (1-16) x US-09-764-869-2259 (1-9745)

Qy 4 G1YrGAspAspAspAspArgArgProSerArgLys 15

Db 6855 GGCAGTGCATGGGACGACGAGCCAGAGAGAA 6820

## RESULT 23

```
US-09-920-923-27
/ Sequence 27, Application US/09920923
/ Publication No. US2003002273A1
/ GENERAL INFORMATION:
/ APPLICANT: Pasamontes, Luis
/ APPLICANT: Tsygankov, Yuri
/ TITLE OF INVENTION: Improved Fermentative Carotenoid Production
/ FILE REFERENCE: Improved Fermentative Carotenoid
/ CURRENT APPLICATION NUMBER: US/09/920,923
/ CURRENT FILING DATE: 2001-08-02
/ PRIOR APPLICATION NUMBER: 08/980,832
/ PRIOR FILING DATE: 1997-12-01
/ NUMBER OF SEQ ID NOS: 66
/ SOFTWARE: PatentIn Ver. 2.1
/ SEQ ID NO 27
/ LENGTH: 11233
/ TYPE: DNA
/ ORGANISM: Unknown
/ FEATURE:
/ OTHER INFORMATION: Description of Unknown Organism: Plasmid pZea4
US-09-920-923-27
```

## Alignment Scores:

```
Pred. No.: 528 Length: 11233
Score: 43.00 Matches: 9
Percent Similarity: 69.23% Conservative: 0
Best Local Similarity: 69.23% Mismatches: 4
Query Match: 47.78% Indels: 0
DB: 9 Gaps: 0
```

US-10-020-441-3 (1-16) x US-09-920-923-27 (1-11233)

Qy 2 G1YrG1YrGAspAspAspAspArgArgProSerArg 14

Db 6700 GGAAGCGGTGGGATCGCGGCGATCGATGCCACGCCG 6738

## RESULT 24

```
US-10-091-504-2258/c
/ Sequence 2258, Application US/10091504
```

```
Publication No. US20030059908A1
GENERAL INFORMATION:
APPLICANT: Rosen et al.
TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
FILE REFERENCE: PC007C1
CURRENT APPLICATION NUMBER: US/10/091,504
CURRENT FILING DATE: 2002-03-07
NUMBER OF SEQ ID NOS: 2442
Prior Application removed - See File Wrapper or Palm
SOFTWARE: Patentln Ver. 2.0
SEQ ID NO 2258
LENGTH: 12149
TYPE: DNA
ORGANISM: Homo sapiens
US-10-091-504-2258

Alignment Scores:
Pred. No.: 572          Length: 12149
Score: 43.00           Matches: 7
Percent Similarity: 75.00% Conservative: 2
Best Local Similarity: 58.33% Mismatches: 3
Query Match: 47.78%    Indels: 0
DB: 9                  Gaps: 0

US-10-020-441-3 (1-16) x US-10-091-504-2258 (1-12149)
Oy 4 G1yArGaAsPaAsnAspArGaSPProSeArGlyS 15
Db 9262 GGCAGTACGATGGGACAGAGAGCCAGAGAGANA 9227

RESULT 25
US-09-764-869-2258/c
Sequence 2258, Application US/09764869
Patent No. US20020061521A1
GENERAL INFORMATION:
APPLICANT: Rosen et al.
TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
FILE REFERENCE: PC007
CURRENT APPLICATION NUMBER: US/09/764,869
CURRENT FILING DATE: 2001-01-17
Prior Application data removed - refer to PALM or file wrapper
NUMBER OF SEQ ID NOS: 2442
SOFTWARE: Patentln Ver. 2.0
SEQ ID NO 2258
LENGTH: 12149
TYPE: DNA
ORGANISM: Homo sapiens
US-09-764-869-2258

Alignment Scores:
Pred. No.: 572          Length: 12149
Score: 43.00           Matches: 7
Percent Similarity: 75.00% Conservative: 2
Best Local Similarity: 58.33% Mismatches: 3
Query Match: 47.78%    Indels: 0
DB: 10                 Gaps: 0

US-10-020-441-3 (1-16) x US-09-764-869-2258 (1-12149)
Oy 4 G1yArGaAsPaAsnAspArGaSPProSeArGlyS 15
Db 9262 GGCAGTACGATGGGACAGAGAGCCAGAGAGANA 9227

RESULT 26
US-09-738-626-1/c
Sequence 1, Application US/09738626
Publication No. US20020197605A1
GENERAL INFORMATION:
APPLICANT: NAKAGAMA, SATOSHI
APPLICANT: MIZOGUCHI, HIROSHI
APPLICANT: ANDO, SEIKO
APPLICANT: HAYASHI, MIKIRO
APPLICANT: OCHIAI, KEIKO
```

```
APPLICANT: YOKOI, HARUHIKO
APPLICANT: TATEISHI, NAOKO
APPLICANT: SENOH, AKIHIRO
APPLICANT: IKEDA, MASATO
APPLICANT: OZAKI, AKIO
TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
FILE REFERENCE: 249-125
CURRENT APPLICATION NUMBER: US/09/738,626
CURRENT FILING DATE: 2000-12-18
Prior Application Number: JP 99/377484
Prior Filing Date: 1999-12-16
Prior Application Number: JP 00/159162
Prior Filing Date: 2000-04-07
Prior Application Number: JP 00/280988
Prior Filing Date: 2000-08-03
NUMBER OF SEQ ID NOS: 7059
SOFTWARE: Patentln ver. 3.0
SEQ ID NO 1
LENGTH: 3309400
TYPE: DNA
ORGANISM: Corynebacterium glutamicum
US-09-738-626-1

Alignment Scores:
Pred. No.: 1,676+05     Length: 3309400
Score: 43.00           Matches: 8
Percent Similarity: 75.00% Conservative: 1
Best Local Similarity: 66.67% Mismatches: 3
Query Match: 47.78%    Indels: 0
DB: 9                  Gaps: 0

US-10-020-441-3 (1-16) x US-09-738-626-1 (1-3309400)
Oy 1 ValG1yTyrg1yArGaAsPaAsnAspArGaSPPro 12
Db 33307 GTGGGGGTTGTGCGTATGAAATGACGCCGATCCC 33272

RESULT 27
US-09-878-574-7832
Sequence 7832, Application US/09878574
Patent No. US20020110548A1
GENERAL INFORMATION:
APPLICANT: Byrum, Joseph R.
APPLICANT: La Rosa, Thomas J.
APPLICANT: Thompson, Michael D.
TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated with
FILE REFERENCE: 38-21(15401)B
CURRENT APPLICATION NUMBER: US/09/878,574
CURRENT FILING DATE: 2001-12-21
Prior Application Number: 09/333,535
Prior Filing Date: 1999-06-14
NUMBER OF SEQ ID NOS: 15775
SEQ ID NO 7832
LENGTH: 261
TYPE: DNA
ORGANISM: Glycine max
OTHER INFORMATION: Clone ID: 701100394H1
US-09-878-574-7832

Alignment Scores:
Pred. No.: 16 9         Length: 261
Score: 42.00           Matches: 7
Percent Similarity: 72.73% Conservative: 1
Best Local Similarity: 63.64% Mismatches: 3
Query Match: 46.67%    Indels: 0
DB: 10                 Gaps: 0

US-10-020-441-3 (1-16) x US-09-878-574-7832 (1-261)
Oy 1 ValG1yTyrg1yArGaAsPaAsnAspArGaSP 11
Db 86 CTCGATACGGCGCGACGACGAAGACGAAGAC 116
```

US-10-020-441-3 (1-16) x US-09-764-864-575 (1-1742)	
Oy	4 G1VArGAspAaPaAspARgAspPrSeArGhYsAa 15
DB	618 GAGCGTGGGACCCAGACCGGAGCCAGTACGATGAGAA 653
RESULT 30	
US-10-047-260-29/c	
Sequence 29, Application US/10047260	
Patent No. US20020164706A1	
GENERAL INFORMATION:	
APPLICANT: Huang, Lisea	
APPLICANT: LaRossa, Robert	
TITLE OF INVENTION: HIGH LEVEL PROMOTERS FROM CYANOBACTERIA	
FILE REFERENCE: C11715 US NA	
CURRENT APPLICATION NUMBER: US/10/047,260	
CURRENT FILING DATE: 2002-01-15	
PRIOR APPLICATION NUMBER: 60/264,925	
PRIOR FILING DATE: 2001-01-30	
NUMBER OF SEQ ID NOS: 40	
SOFTWARE: Microsoft Office 97	
SEQ ID NO 29	
LENGTH: 1851	
TYPE: DNA	
ORGANISM: Synechocystis sp. strain PCC6803	
FEATURE:	
NAME/KEY: CDS	
LOCATION: (1)..(1851)	
US-10-047-260-29	
Alignment Scores:	
Pred. No.: 126	
Score: 42.00	
Percent Similarity: 81.82%	
Best Local Similarity: 72.73%	
Query Match: 46.67%	
DB: 9	
US-10-020-441-3 (1-16) x US-10-047-260-29 (1-1851)	
Oy	6 AsPaAspAaPaAspARgAspPrSeArGhYsAa 16
DB	970 GATCGAAGACGACCGGAGCGCATCAAGCGAGAT 938
RESULT 31	
US-10-047-260-35/c	
Sequence 35, Application US/10047260	
Patent No. US20020164706A1	
GENERAL INFORMATION:	
APPLICANT: Huang, Lisea	
APPLICANT: McCluskey, Michael	
TITLE OF INVENTION: HIGH LEVEL PROMOTERS FROM CYANOBACTERIA	
FILE REFERENCE: C11715 US NA	
CURRENT APPLICATION NUMBER: US/10/047,260	
CURRENT FILING DATE: 2002-01-15	
PRIOR APPLICATION NUMBER: 60/264,925	
PRIOR FILING DATE: 2001-01-30	
NUMBER OF SEQ ID NOS: 40	
SOFTWARE: Microsoft Office 97	
SEQ ID NO 35	
LENGTH: 1884	
TYPE: DNA	
ORGANISM: Synechocystis sp. strain PCC6803	
FEATURE:	
NAME/KEY: CDS	
LOCATION: (1)..(1884)	



US-10-047-260-35

## Alignment Scores:

Pred. No.:	128	Length:	1884
Score:	42.00	Matches:	8
Percent Similarity:	81.82%	Conservative:	1
Best Local Similarity:	72.73%	Mismatches:	2
Query Match:	46.67%	Indels:	0
DB:	9	Gaps:	0

US-10-020-441-3 (1-16) x US-10-047-260-35 (1-1884)

Qy 6 AspaPaanaPaArgaPProSeArGLysAsn 16

Db 1003 GATCGAAGCAGCCGGCAGCATCAAGCGAGAAT 971

## RESULT 32

US-09-752-983-1

Sequence 1, Application US/09752983

Patent No. US20010016575A1

GENERAL INFORMATION:

APPLICANT: Loren J. Miraglia, Pamela Nero, Mark J.

APPLICANT: Graham, Brett P. Monia

TITLE OF INVENTION: ANTISENSE MODULATION OF HUMAN KDM2

TITLE OF INVENTION: EXPRESSION

NUMBER OF SEQUENCES: 271

CORRESPONDENCE ADDRESS:

ADDRESSEE: Law Offices of Jane Massey Licata

STREET: 66 East Main Street

CITY: Marlton

STATE: NJ

COUNTRY: U.S.A.

ZIP: 08053

COMPUTER READABLE FORM:

MEDIUM TYPE: DISKETTE, 3.5 INCH, 1.44 MB STORAGE

COMPUTER: IBM PC

OPERATING SYSTEM: WINDOWS 95

SOFTWARE: WORDPERFECT 6.0

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/752,983

FILING DATE: 02-Jan-2001

CLASSIFICATION:

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 09/280,805

FILING DATE: &lt;Unknown&gt;

ATTORNEY/AGENT INFORMATION:

NAME: Licata, Jane Massey

REGISTRATION NUMBER: 32,257

REFERENCE/DOCKET NUMBER: ISPH-0346

TELECOMMUNICATION INFORMATION:

TELEPHONE: 609-810-1515

TELEFAX: 609-810-1454

INFORMATION FOR SEQ ID NO: 1:

SEQUENCE CHARACTERISTICS:

LENGTH: 2372 base pairs

TYPE: Nucleic Acid

STRANDEDNESS: Single

TOPOLOGY: Unknown

ANTI-SENSE: No

PUBLICATION INFORMATION:

AUTHORS: Oliner, J.D.

AUTHORS: Kinzler, K.W.

AUTHORS: Meltzer, P.S.

AUTHORS: George, D.L.

AUTHORS: Vogelstein, B.

TITLE: Amplification of a gene encoding a

JOURNAL: Nature

VOLUME: 358

ISSUE: 6381

PAGES: 80-83

DATE: 02-JUL-1992

US-09-752-983-1

## Alignment Scores:

Pred. No.:	162	Length:	2372
Score:	42.00	Matches:	7
Percent Similarity:	90.00%	Conservative:	2
Best Local Similarity:	70.00%	Mismatches:	1
Query Match:	46.67%	Indels:	0
DB:	10	Gaps:	0

US-10-020-441-3 (1-16) x US-09-752-983-1 (1-2372)

Qy 4 GLyArGaPaPaanaPaArgaPProSeAr 13

Db 84 GGCGCGAGCCCTCTGACCGAGATCTGCT 113

## RESULT 33

US-09-956-425-7

Sequence 7, Application US/09956425

Patent No. US20020045192A1

GENERAL INFORMATION:

APPLICANT: Kriwacki, Richard

APPLICANT: Botner, Brian

APPLICANT: Lewis, William

TITLE OF INVENTION: Arf and Hdm2 Interaction Domains and Method of Use Thereof

FILE REFERENCE: 1340/1/035

CURRENT APPLICATION NUMBER: US/09/956,425

CURRENT FILING DATE: 2001-09-19

NUMBER OF SEQ ID NOS: 25

SOFTWARE: PatentIn version 3.1

SEQ ID NO: 7

LENGTH: 2372

TYPE: DNA

ORGANISM: Homo sapiens

US-09-956-425-7

## Alignment Scores:

Pred. No.:	162	Length:	2372
Score:	42.00	Matches:	7
Percent Similarity:	90.00%	Conservative:	2
Best Local Similarity:	70.00%	Mismatches:	1
Query Match:	46.67%	Indels:	0
DB:	10	Gaps:	0

US-10-020-441-3 (1-16) x US-09-956-425-7 (1-2372)

Qy 4 GLyArGaPaPaanaPaArgaPProSeAr 13

Db 84 GGCGCGAGCCCTCTGACCGAGATCTGCT 113

## RESULT 34

US-09-851-771A-1

Sequence 1, Application US/09851771A

Patent No. US20020151511A1

GENERAL INFORMATION:

APPLICANT: Loren J. Miraglia, Pamela Nero, Mark J.

APPLICANT: Graham, Brett P. Monia

TITLE OF INVENTION: ANTISENSE OLIGONUCLEOTIDE

MODULATION OF HUMAN KDM2 EXPRESSION

NUMBER OF SEQUENCES: 32

CORRESPONDENCE ADDRESS:

ADDRESSEE: Law Offices of Jane Massey Licata

STREET: 66 East Main Street

CITY: Marlton

STATE: NJ

COUNTRY: U.S.A.

ZIP: 08053

COMPUTER READABLE FORM:

MEDIUM TYPE: DISKETTE, 3.5 INCH, 1.44 MB STORAGE

COMPUTER: IBM 486

OPERATING SYSTEM: WINDOWS FOR WORKGROUPS

SOFTWARE: WORDPERFECT 5.1

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/851,771A

```

; FILING DATE: 09-May-2001
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 09/048,810
; FILING DATE: 1998-03-26
; ATTORNEY/AGENT INFORMATION:
; NAME: Licata, Jane Massey
; REGISTRATION NUMBER: 32,257
; REFERENCE/DOCKET NUMBER: ISPH-0302
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 609-779-2400
; TELEFAX: 609-810-1454
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2372 base pairs
; TYPE: Nucleic Acid
; STRANDEDNESS: Single
; TOPOLOGY: Unknown
; ANTI-SENSE: NO
; PUBLICATION INFORMATION:
; AUTHORS: Oliner,J.D.,Kinzler,K.W.,Meltzer,P.S.,George,D.L.,Vogelstein,B.
; TITLE: Amplification of a gene encoding a p53-associated protein in hum
; JOURNAL: Nature
; VOLUME: 358
; ISSUE: 6381
; PAGES: 80-83
; DATE: 02-JUL-1992
; SEQUENCE DESCRIPTION: SEQ ID NO: 1:
US-09-851-771A-1

Alignment Scores:
Pred. No.: 162 Length: 2372
Score: 42.00 Matches: 7
Percent Similarity: 90.00% Conservative: 2
Best Local Similarity: 70.00% Mismatches: 1
Query Match: 46.67% Indels: 0
DB: 10 Gaps: 0

US-10-020-441-3 (1-16) x US-09-851-771A-1 (1-2372)

Cy 4 G|Y|A|G|A|P|A|S|P|A|S|P|A|G|S|P|P|O|S|E|R| 13
Db 84 G|G|C|G|G|A|C|C|C|T|G|A|C|C|G|A|G|A|T|C|T|G|C|T| 113

RESULT 35
US-09-925-301-493
; Sequence 493, Application US/09925301
; Patent No. US20020052308A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
; FILE REFERENCE: PA106
; CURRENT APPLICATION NUMBER: US/09/925,301
; PRIOR FILING DATE: 2001-08-10
; PRIOR APPLICATION NUMBER: PCT/US00/05882
; PRIOR FILING DATE: 2000-03-08
; PRIOR APPLICATION NUMBER: 60/124,270
; PRIOR FILING DATE: 1999-03-12
; NUMBER OF SEQ ID NOS: 1694
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 493
; LENGTH: 4108
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-925-301-493

Alignment Scores:
Pred. No.: 285 Length: 4108
Score: 42.00 Matches: 6
Percent Similarity: 90.91% Conservative: 4
Best Local Similarity: 54.55% Mismatches: 1
Query Match: 46.67% Indels: 0
DB: 10 Gaps: 0
```

```

US-10-020-441-3 (1-16) x US-09-925-301-493 (1-4108)

Cy 5 A|G|A|P|A|S|P|A|S|P|A|G|S|P|P|O|S|E|R|A|G|S|A|N| 15
Db 2517 C|G|G|A|G|G|A|G|A|T|G|A|C|C|C|T|G|A|G|A|G|A| 2549

RESULT 36
US-10-101-464A-165/c
; Sequence 165, Application US/10101464A
; Publication No. US20030046728A1
; GENERAL INFORMATION:
; APPLICANT: Strabala, Timothy
; APPLICANT: Nieuwenhuizen, Nicolaas
; APPLICANT: Higgins, Colleen M.
; TITLE OF INVENTION: Compositions isolated from Plant Cells
; TITLE OF INVENTION: and their Use in the Modification of Plant Cell Signaling
; FILE REFERENCE: 11000.1020c2
; CURRENT APPLICATION NUMBER: US/10/101,464A
; PRIOR FILING DATE: 2002-03-18
; PRIOR APPLICATION NUMBER: 09/704,302
; PRIOR FILING DATE: 2000-11-01
; PRIOR APPLICATION NUMBER: 09/228,986
; PRIOR FILING DATE: 1999-01-12
; PRIOR APPLICATION NUMBER: 60/162,866
; PRIOR FILING DATE: 1999-11-01
; PRIOR APPLICATION NUMBER: PCT/US00/00724
; PRIOR FILING DATE: 2000-01-11
; NUMBER OF SEQ ID NOS: 989
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 165
; LENGTH: 438
; TYPE: DNA
; ORGANISM: Eucalyptus grandis
US-10-101-464A-165

Alignment Scores:
Pred. No.: 43.4 Length: 438
Score: 41.00 Matches: 8
Percent Similarity: 73.33% Conservative: 3
Best Local Similarity: 53.33% Mismatches: 4
Query Match: 45.56% Indels: 0
DB: 9 Gaps: 0

US-10-020-441-3 (1-16) x US-10-101-464A-165 (1-438)

Cy 2 G|Y|T|Y|G|Y|A|G|A|P|A|S|P|A|S|P|A|G|S|P|P|O|S|E|R|A|G|S|A|N| 16
Db 397 G|G|T|A|T|T|G|A|G|A|C|T|A|T|C|G|A|C|C|G|A|G|A|C|T|G|C|G|C|G|A|G|A|G|C| 353

RESULT 37
US-09-864-761-11028/c
; Sequence 11028, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharon G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
; TITLE OF INVENTION: GENE EXPRESSION ANALYSIS BY MICROARRAY
; FILE REFERENCE: Aeomica-X-1
; CURRENT APPLICATION NUMBER: US/09/864,761
; PRIOR FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/632,356
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: GB 24263,6
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/236,359
```

```

; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US01/00666
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00664
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00669
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00665
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00668
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00663
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00662
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00661
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00670
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: US 09/608,408
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/774,203
; PRIOR FILING DATE: 2001-01-29
; NUMBER OF SEQ ID NOS: 49117
; SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 11028
; LENGTH: 462
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AC005866.3
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 0.51
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 0.57
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 0.53
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 0.64
US-09-864-761-11028

Alignment Scores:
Pred. No.: 45.8 Length: 462
Score: 41.00 Matches: 8
Percent Similarity: 69.23% Conservative: 1
Best Local Similarity: 61.54% Mismatches: 4
Query Match: 45.56% Indels: 0
DB: 10 Gaps: 0

US-10-020-441-3 (1-16) x US-09-864-761-11028 (1-462)
Qy 3 TTTGGAAAGATTAAGAAAGATCGACGACGCGTAAAGAAAT 15
Db 294 TACGGAGAGGGGACGAGAGGAGACAGTCAAGCGCAG 256

RESULT 38
US-09-764-868-234/C
; Sequence 234, Application US/09764868
; Patent No. US2002016871A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PT232
; CURRENT APPLICATION NUMBER: US/09/764,868
; PRIOR FILING DATE: 2001-01-17
; Prior application data removed - refer to PALM or file wrapper
; NUMBER OF SEQ ID NOS: 1510
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 234
; LENGTH: 861
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-764-868-234
```

```

Alignment Scores:
Pred. No.: 86.8 Length: 861
Score: 41.00 Matches: 7
Percent Similarity: 71.43% Conservative: 3
Best Local Similarity: 50.00% Mismatches: 4
Query Match: 45.56% Indels: 0
DB: 9 Gaps: 0

US-10-020-441-3 (1-16) x US-09-764-868-234 (1-861)
Qy 3 TTTGGAAAGATTAAGAAAGATCGACGACGCGTAAAGAAAT 16
Db 594 TTTGGAAAGATTAAGAAAGATCGACGACGCGTAAAGAAAT 553

RESULT 39
US-09-974-300-699/C
; Sequence 699, Application US/09974300
; Patent No. US20020146721A1
; GENERAL INFORMATION:
; APPLICANT: Berka, Randy M.
; TITLE OF INVENTION: Methods For Monitoring Multiple Gene
; FILE REFERENCE: 10085,500-US
; CURRENT APPLICATION NUMBER: US/09/974,300
; PRIOR FILING DATE: 2001-10-05
; PRIOR APPLICATION NUMBER: 09/680,598
; PRIOR FILING DATE: 2000-10-06
; PRIOR APPLICATION NUMBER: 60/279,526
; PRIOR FILING DATE: 2001-03-27
; NUMBER OF SEQ ID NOS: 8481
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 699
; LENGTH: 1209
; TYPE: DNA
; ORGANISM: Bacillus licheniformis
US-09-974-300-699

Alignment Scores:
Pred. No.: 123 Length: 1209
Score: 41.00 Matches: 7
Percent Similarity: 68.75% Conservative: 4
Best Local Similarity: 43.75% Mismatches: 5
Query Match: 45.56% Indels: 0
DB: 10 Gaps: 0

US-10-020-441-3 (1-16) x US-09-974-300-699 (1-1209)
Qy 1 ValGlyTyrGlyArgAspAsnAspArgAspProSerArgLysAsn 16
Db 256 ATCAATGTCCTCCGATGATGACGCGCATCCCAATCGCTAAAT 209

RESULT 40
US-10-125-540-24
; Sequence 24, Application US/10125540
; Publication No. US20030059875A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PT214C1
; CURRENT APPLICATION NUMBER: US/10/125,540
; PRIOR FILING DATE: 2002-04-19
; Prior application removed - See file wrapper or Palm
; NUMBER OF SEQ ID NOS: 646
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 24
; LENGTH: 3027
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-125-540-24

Alignment Scores:
```

Pred. No.:	315	Length:	3027
Score:	41.00	Matches:	7
Percent Similarity:	88.89%	Conservative:	1
Best Local Similarity:	77.78%	Mismatches:	1
Query Match:	45.56%	Indels:	0
DB:	9	Gaps:	0

US-10-020-441-3 (1-16) x US-10-125-540-24 (1-3027)

Qy	3	TyrGIYArgASPASpASpArgASP	11
Db	2444	TATGGGGGTGATGATAGTGACCGAGAC	2470

Search completed: April 8, 2003, 20:55:36  
 Job time : 278.377 secs

GenCore version 5.1.4 p5\_4578  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: Apr11 8, 2003, 18:07:04 / Search time 222.557 Seconds

(without alignments)  
1164.319 Million cell updates/sec

Title: US-10-020-441-3

Perfect score: 90

Sequence: 1 VGYGRDDNDPDRSKN 16

Scoring table: BIOSUM62

Xgapop 10.0, Ygapext 0.5

Fgapop 6.0, Fgapext 7.0

Delop 6.0, Delext 7.0

Searched: 16154066 seqs, 8097743376 residues

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+ p2n.model -DEV=xlh

-Q=/cgm2\_1/USPTO.spool/US10020441/runat\_02042003\_093012\_20981/app.query.fasta.1.462

-DB=EST -QFMT=fasta1 -SUFIX=1 -MINMATCH=0.1 -LOOPEXT=0

-UNITS=bits -START=1 -END=1 -MATRIX=biolum62 -TRANS=human40.cdi -LIST=45

-DOCALLIGN=200 -THR SCORE=pcr -THR MAX=100 -THR MIN=0 -ALIGN=40 -MODE=LOCAL

-OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000

-USER=US10020441.CCGN\_1.1\_2013.@runat\_02042003\_093012\_20981 -NCPU=6 -ICPU=3

-NO\_XLPHY -NO\_MMAPP -LARGQUERY -NEG SCORES=0 -WAIT -LONGLOG -DEV TIMEOUT=120

-WARR TIMEOUT=30 -THRADG=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7

-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

EST: \*

1: em\_estba:\*

2: em\_esthum:\*

3: em\_estin:\*

4: em\_estmu:\*

5: em\_estov:\*

6: em\_estpl:\*

7: em\_estro:\*

8: em\_hlc:\*

9: gb\_est1:\*

10: gb\_est2:\*

11: gb\_hlc:\*

12: gb\_est3:\*

13: gb\_est4:\*

14: gb\_est5:\*

15: em\_estfun:\*

16: em\_estom:\*

17: gb\_gss:\*

18: em\_gss\_hum:\*

19: em\_gss\_inv:\*

20: em\_gss\_pln:\*

21: em\_gss\_vit:\*

22: em\_gss\_fun:\*

23: em\_gss\_mam:\*

24: em\_gss\_mus:\*

25: em\_gss\_other:\*

26: em\_gss\_pro:\*

27: em\_gss\_rod:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	52	57.8	788	9	AU131395
2	52	57.8	814	12	BE870296
3	52	57.8	845	9	AU131025
4	52	57.8	1017	14	BQ878109
5	52	57.8	1218	13	BM462474
6	50.5	56.1	955	12	BF301612
7	50	55.6	407	10	AV818818
8	50	55.6	574	13	BI546228
9	49	54.4	295	10	BI819948
10	49	54.4	495	17	BH829275
11	49	54.4	514	10	AV600039
12	49	54.4	549	13	BJ127850
13	49	54.4	556	10	AV606897
14	49	54.4	668	10	BB642648
15	49	54.4	682	10	BB642873
16	49	54.4	737	17	BH839335
17	48.5	53.9	369	10	AW414724
18	48.5	53.9	376	17	AZ923074
19	48	53.3	372	17	CNS03029
20	48	53.3	461	9	AL376173
21	48	53.3	477	9	AL379254
22	48	53.3	500	17	AQ964575
23	48	53.3	521	17	AQ964574
24	48	53.3	525	9	AL379255
25	48	53.3	560	13	BI920718
26	48	53.3	573	10	AW679964
27	48	53.3	713	13	BM108979
28	48	53.3	919	9	AU176144
29	48	53.3	919	17	AG083636
30	48	53.3	980	14	BO704202
31	48	53.3	1073	17	CNS061C1
32	47.5	52.8	446	9	AU184369
33	47.5	52.8	447	9	AU029604
34	47.5	52.8	449	9	AU184633
35	47.5	52.8	467	14	BO906523
36	47.5	52.8	480	13	BI809420
37	47.5	52.8	497	13	BI306405
38	47.5	52.8	531	13	BM419755
39	47.5	52.8	713	9	AU082192
40	47	52.2	275	9	AU009417
41	47	52.2	325	10	AW727714
42	47	52.2	355	10	BE366885
43	47	52.2	368	13	BJ221203
44	47	52.2	483	10	BE551439
45	47	52.2	503	12	BG515602

#### ALIGNMENTS

RESULT 1	LOCUS	DEFINITION	ACCESSION	VERSION	KEYWORDS	SOURCE	ORGANISM	REFERENCE	AUTHORS
AU131395/c	AU131395	AU131395 NT2RP3 Homo sapiens CDNA clone NT2RP3002499 5', mRNA	AU131395	AU131395	GI:10991749	human.	Homo sapiens	1 (bases 1 to 788)	Ota, T., Nishikawa, T., Suzuki, Y., Ishii, S., Saito, K., Kawai, Y., Yamamoto, J., Wakamatsu, A., Nakamura, Y., Nagai, T., Sugano, S. and

TITLE HRI human cDNA project  
JOURNAL Unpublished (2000)  
COMMENT Contact: Takao Isogai  
Genomics Laboratory  
Helix Research Institute  
1532-3 Yana, Kisarazu, Chiba 292-0812, Japan  
Tel: 81-438-52-3975  
Fax: 81-438-52-3986  
Email: genomics@hri.co.jp

HRI human cDNA project: 5'- & 3'-end one pass sequencing: Helix  
Research Institute; cDNA library construction: Department of  
Virology, Institute of Medical Science, University of Tokyo, and  
Helix Research Institute.  
Location/Qualifiers

FEATURES  
source 1..788  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="NT2RP3002499"  
/clone\_lib="NT2RP3"  
/cell\_type="teratocarcinoma"  
/note="vector: pME18SF13; mRNA from NT2 neuronal precursor  
cells after 2-weeks retinoic acid (RA) induction"

BASE COUNT 244 a 164 c 167 g 210 t 3 others  
ORIGIN

## Alignment Scores:

Pred. No.: 87.2 Length: 788  
Score: 52.00 Matches: 11  
Percent Similarity: 85.71% Conservative: 1  
Best Local Similarity: 78.57% Mismatches: 1  
Query Match: 57.78% Indels: 1  
DB: 9 Gaps: 0

US-10-020-441-3 (1-16) x AU131395 (1-788)

Qy 1 ValGlyTYrGlyArgAspAspAsnAsp-ArgAspProSer 13  
Db 299 ATGGGATATGGTACTGACGACACATGATCAGAGACCCCAAGT 260

RESULT 2 BE870296 814 bp mRNA linear EST 20-OCT-2000  
BE870296/c LOCUS 601447441F1 NIH\_MGC\_65 Homo sapiens cDNA clone IMAGE:3851410 5',  
DEFINITION mRNA sequence.  
ACCESSION BE870296  
VERSION BE870296.1 GI:10319072  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
1 (bases 1 to 814)  
NIH-MGC http://mgc.nci.nih.gov/.  
National Institutes of Health, Mammalian Gene Collection (MGC)  
Unpublished (1999)  
Contact: Robert Strausberg, Ph.D.  
Email: cgapbs-remail.nih.gov  
Tissue Procurement: ATCC  
Library Preparation: Life Technologies, Inc.  
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (ILNL)  
DNA Sequencing by: Incyte Genomics, Inc.  
Clone distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/ILNL at:  
http://image.llnl.gov  
Plate: L1AM9572 row: a column: 11  
High quality sequence stop: 605.  
Location/Qualifiers

FEATURES  
source 1..814  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:3851410"

/clone\_lib="NIH\_MGC\_65"  
/tissue\_type="adenocarcinoma"  
/lab\_host="DH10B (phage-resistant)"  
/note="Organ: colon; Vector: PCMV-SPORE; Site 1: NotI;  
Site 2: SalI; Cloned unidirectionally. Primer: Oligo dT.  
Average insert size 1.8 kb. Library constructed by Life  
Technologies." Location/Qualifiers

BASE COUNT 254 a 213 c 180 g 167 t  
ORIGIN

## Alignment Scores:

Pred. No.: 90.7 Length: 814  
Score: 52.00 Matches: 11  
Percent Similarity: 85.71% Conservative: 1  
Best Local Similarity: 78.57% Mismatches: 1  
Query Match: 57.78% Indels: 1  
DB: 12 Gaps: 0

US-10-020-441-3 (1-16) x BE870296 (1-814)

Qy 1 ValGlyTYrGlyArgAspAspAsnAsp-ArgAspProSer 13  
Db 185 ATGGGATATGGTACTGACGACACATGATCAGAGACCCCAAGT 146

RESULT 3 AU131025 845 bp mRNA linear EST 01-AUG-2002  
AU131025/c LOCUS AU131025 NT2RP3 Homo sapiens cDNA clone NT2RP3001871 5', mRNA  
DEFINITION sequence.  
ACCESSION AU131025  
VERSION AU131025.1 GI:10991379  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
1 (bases 1 to 845)  
Ota, T., Nishikawa, T., Suzuki, Y., Ishii, S., Saito, K., Kawai, Y.,  
Yamamoto, J., Wakamatsu, A., Nakamura, Y., Nagai, T., Sugano, S. and  
Isogai, T.  
HRI human cDNA project  
Unpublished (2000)  
Contact: Takao Isogai  
Genomics Laboratory  
Helix Research Institute  
1532-3 Yana, Kisarazu, Chiba 292-0812, Japan  
Tel: 81-438-52-3975  
Fax: 81-438-52-3986  
Email: genomics@hri.co.jp  
HRI human cDNA project: 5'- & 3'-end one pass sequencing: Helix  
Research Institute; cDNA library construction: Department of  
Virology, Institute of Medical Science, University of Tokyo, and  
Helix Research Institute.  
Location/Qualifiers

FEATURES  
source 1..845  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="NT2RP3001871"  
/clone\_lib="NT2RP3"  
/cell\_type="teratocarcinoma"  
/note="vector: pME18SF13; mRNA from NT2 neuronal precursor  
cells after 2-weeks retinoic acid (RA) induction"

BASE COUNT 257 a 167 c 183 g 235 t 3 others  
ORIGIN

## Alignment Scores:

Pred. No.: 95 Length: 845  
Score: 52.00 Matches: 11  
Percent Similarity: 85.71% Conservative: 1  
Best Local Similarity: 78.57% Mismatches: 1  
Query Match: 57.78% Indels: 1  
DB: 9 Gaps: 0

US-10-020-441-3 (1-16) x AUJ1025 (1-845)

OY 1 ValGlyTYrGlyAArgAspAspAspAsp-ArgAspProSer 13  
Db 365 ATGGGATATGGTACTGACGACAAATGATCAGAGACCAAGT 326

RESULT 4 BQ878109/c 1017 bp mRNA linear EST 16-AUG-2002

LOCUS BQ878109 AGENCOURT 8152145 lupski\_dorsal\_root\_ganglion Homo sapiens CDNA  
DEFINITION clone IMAGE:6184030 5', mRNA sequence.

ACCESSION BQ878109  
VERSION BQ878109.1 GI:22270117  
KEYWORDS EST.

SOURCE human.

ORGANISM Homo sapiens

REFERENCE Bukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
1 (bases 1 to 1017)

AUTHORS NIH-MGC http://mgc.nci.nih.gov/.

TITLE National Institutes of Health, Mammalian Gene Collection (MGC)

JOURNAL Unpublished (1999)

COMMENT Contact: Robert Strausberg, Ph.D.  
Email: cgabs-remail.nih.gov

Tissue Procurement: Dr. James R. Lupski

CDNA Library Preparation: Life Technologies, Inc.

CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNLN)

DNA Sequencing by: Agencourt Bioscience Corporation

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LNLN at:

http://image.lnl.gov

Plate: LHAM13572 row: 1 column: 23

High quality sequence stop: 322.

FEATURES

source

Location/Qualifiers

1..1017

/organism="Homo sapiens"

/db\_xref="taxon:9606"

/clone="IMAGE:6184030"

/clone\_lib="lupski\_dorsal\_root\_ganglion"

/sex="male"

/tissue\_type="dorsal root ganglia"

/dev\_stage="adult, 36 yr"

/lab\_host="DH10B"

/note="Vector: pCMV-SPORT6 (Life Technologies); Site\_1: NotI, Site\_2: SalI, CDNA made by oligo-dt priming.

Directionally cloned using the following adaptors:

5'-TCGACCCACGGGCTCG-3' and

5'-GACTGATTTGATCGGAGCGCGCCCT(15)-3', size selected >

1 kb for average insert length 1.7 kb. This is a primary

library, non-amplified, library constructed by life

Technologies and donated by J. Lupski, M.D./Ph.D. (Baylor

College of Medicine) and is available through life

Technologies."

BASE COUNT 292 a 269 c 251 g 205 t

ORIGIN

Alignment Scores:

Pred. No.: 120 Length: 1017

Score: 52.00 Matches: 11

Percent Similarity: 85.71% Conservative: 1

Best Local Similarity: 78.57% Mismatches: 1

Query Match: 57.78% Indels: 1

DB: 14 Gaps: 0

US-10-020-441-3 (1-16) x BQ878109 (1-1017)

OY 1 ValGlyTYrGlyAArgAspAspAsp-ArgAspProSer 13

Db 90 ATGGGATATGGTACTGACGACAAATGATCAGAGACCAAGT 51

RESULT 5 BM462474

LOCUS BM462474 1218 bp mRNA linear EST 05-FEB-2002

DEFINITION AGENCOURT 6426330 NIH\_MGC\_71 Homo sapiens CDNA clone IMAGE:5518144

ACCESSION BM462474

VERSION BM462474.1 GI:18511514

KEYWORDS EST.

SOURCE human.

ORGANISM Homo sapiens

REFERENCE Bukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
1 (bases 1 to 1218)

AUTHORS NIH-MGC http://mgc.nci.nih.gov/.

TITLE National Institutes of Health, Mammalian Gene Collection (MGC)

JOURNAL Unpublished (1999)

COMMENT Contact: Robert Strausberg, Ph.D.  
Email: cgabs-remail.nih.gov

Tissue Procurement: ATCC

CDNA Library Preparation: Life Technologies, Inc.

CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNLN)

DNA Sequencing by: Agencourt Bioscience Corporation

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LNLN at:

http://image.lnl.gov

Plate: LHAM12178 row: h column: 17

High quality sequence start: 66

High quality sequence stop: 602.

FEATURES

source

Location/Qualifiers

1..1218

/organism="Homo sapiens"

/db\_xref="taxon:9606"

/clone="IMAGE:5518144"

/clone\_lib="NIH MGC 71"

/tissue\_type="ligomyosarcoma"

/lab\_host="DH10B (phage-resistant)"

/note="Organ: uterus; Vector: pCMV-SPORT6, Site\_1: NotI, Site\_2: SalI; Cloned unidirectionally. Primer: Oligo dt.

Average insert size 2.1 kb. "

BASE COUNT 298 a 428 c 268 g 219 t 5 others

ORIGIN

Alignment Scores:

Pred. No.: 149 Length: 1218

Score: 52.00 Matches: 10

Percent Similarity: 78.57% Conservative: 3

Best Local Similarity: 71.43% Mismatches: 0

Query Match: 57.78% Indels: 0

DB: 13 Gaps: 0

US-10-020-441-3 (1-16) x BM462474 (1-1218)

OY 2 GlyTYrGlyAArgAspAspAsp-ArgAspProSerArglys 15

Db 683 GGGTATGGGGGTGGGAAATCCCGGAGCCCTCGCGGAAA 724

RESULT 6 BF301612

LOCUS BF301612 955 bp mRNA linear EST 21-NOV-2000

DEFINITION 60202993F1 NCI\_CGAP\_SG2 Mus musculus CDNA clone IMAGE:4184982 5',

ACCESSION BF301612

VERSION BF301612.1 GI:11248044

KEYWORDS EST.

SOURCE house mouse.

ORGANISM Mus musculus

REFERENCE Bukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
1 (bases 1 to 955)

AUTHORS NIH-MGC http://mgc.nci.nih.gov/.

TITLE National Institutes of Health, Mammalian Gene Collection (MGC)

JOURNAL Unpublished (1999)

COMMENT Contact: Robert Strausberg, Ph.D.  
Email: cgabs-remail.nih.gov

Tissue Procurement: Jeffrey E. Green, M.D.





ALIGNMENT SCORES:

Pred. No.:	128	Length:	574
Score:	50.00	Matches:	9
Percent Similarity:	76.92%	Conservative:	1
Best Local Similarity:	69.23%	Mismatch:	3
Query Match:	55.56%	Indels:	0
DB:	13	Gaps:	0

US-10-020-441-3 (1-16) x B1546298 (1-574)

0y 1 ValGlyTYrGLyArGAspAspAsnAspArgAspProSer 13

Db 238 GTGGGAGCGGCGGTGATGACCGGTGACGACGACCAAC 276

RESULT 9  
B18189485/c

LOCUS  
B18189485/c

DEFINITION  
B18189485 295 bp mRNA linear EST 30-JUN-2000  
musculus cDNA A330054K13, mRNA sequence.

ACCESSION  
B18189485

VERSION  
B18189485.1 GI:8850056

KEYWORDS  
EST.

SOURCE  
house mouse.  
Mus musculus  
Eutheria; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
1 (bases 1 to 295)

REFERENCE  
Kono, H., Aizawa, K., Akahira, S., Akiyama, J., Arakawa, T., Carninci, P., Endo, T., Fukuda, S., Fukunishi, Y., Hara, A., Hayatsu, N., Hirozane, T., Hori, F., Ishii, Y., Ishikawa, J., Ishikawa, T., Itoh, M., Izawa, M., Kadota, K., Kagawa, I., Kai, C., Kawai, J., Kikuchi, N., Kiyosawa, H., Kojima, Y., Kondo, S., Koya, S., Kurihara, C., Kusakabe, M., Matsuyama, T., Miki, R., Mizuno, Y., Nakamura, M., Oda, H., Okazaki, Y., Ono, T., Owa, C., Saito, H., Sakai, C., Sato, K., Shibata, K., Shibata, Y., Shigemoto, Y., Shinagawa, A., Shiraki, T., Sogabe, Y., Sugihara, Y., Suzuki, H., Suzuki, H., Tagawa, A., Takahashi, F., Tomioka, N., Toya, T., Tsunoda, Y., Watabiki, A., Watanabe, S., Yamamura, T., Yamataka, I., Yano, R., Yasunishi, A., Yokota, T., Yoshida, K., Yoshiki, A., Yoshino, M., Muramatsu, M. and Hayashizaki, Y.

TITLE  
RIKEN Mouse ESTs (kono, H., et al.)

JOURNAL  
Unpublished (2000)

COMMENT  
Contact: Yoshinide Hayashizaki  
Laboratory for Genome Exploration Research Group, RIKEN Genomic  
Sciences Center(GSC), Yokohama Institute  
The Institute of Physical and Chemical Research (RIKEN)  
1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan  
Tel: 81-45-503-9222  
Fax: 81-45-503-9216  
Email: genome-res@gsc.riken.go.jp,  
URL: http://genome.gsc.riken.go.jp/  
Carninci, P., Nishiyama, Y., Westover, A., Itoh, M., Nagakura, S., Sasakihara, N., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.  
Thermotranscription and thermotranscription of thermolabile enzymes by  
trehalose and its application for the synthesis of full length  
cDNA. Proc. Natl. Acad. Sci. U.S.A. 95 (2), 520-524 (1998)  
Itoh, M., Kikuchi, T., Akiyama, Y., Shibata, K., Izawa, M., Kawai, J.,  
Tomaru, Y., Carninci, P., Shibata, Y., Ozawa, Y., Muramatsu, M., Okazaki, Y. and Hayashizaki, Y.  
Automated filtration-based high-throughput plasmid preparation  
system. Genome Res. 9 (5), 463-470 (1999)  
Carninci, P. and Hayashizaki, Y.  
High-efficiency full-length cDNA cloning. Methods Enzymol. 303,  
19-44 (1999)  
Please visit our web site (http://genome.etc.riken.go.jp) for  
further details.

FEATURES  
source  
1. 295  
/organism="Mus musculus"  
/db\_xref="taxon:10090"  
/clone\_id="A330054K13"  
/clone\_lib="RIKEN full-length enriched, adult male spinal

[illegible]

```

ORIGIN
Alignment Scores:
Score:      158          Length:      495
Pred. No.:   49.00       Matches:      8
Percent Similarity: 73.33% Conservative: 3
Best Local Similarity: 53.33% Mismatches:  4
Query Match:  54.44%     Indels:    0
DB:           17         Gaps:      0

US-10-020-441-3 (1-16) x BH829275 (1-495)

Oy      2  GLYTYRGIYARGASPASnaPrgAsppProSerARgLyAsn 16
Db      275 GGCAATGGAGGAGGATGAACATGATGAACACCAGAGAGAAATCG 319

RESULT 11
LOCUS      AV600039                    514 bp      mRNA      linear      EST 27-NOV-2001
DEFINITION AV600039 Bos taurus cartilage fetus Bos taurus cDNA clone
            E1C047F06 5', mRNA sequence.
ACCESSION  AV600039
VERSION     AV600039.1 GI:9718784
KEYWORDS    EST.
SOURCE      cow.
ORGANISM    Bos taurus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
            Bovidae; Bovinae; Bos.
            1 (bases 1 to 514)
            Takasuga,A., Hirotsune,S., Itoh,R., Jitohzono,A., Suzuki,H., Aso,H.
            and Sugimoto,Y.
            Establishment of a high throughput EST sequencing system using
            poly(A) tail-removed cDNA libraries and determination of 36,000
            bovine ESTs
            Nucleic Acids Res. 29 (22), E108 (2001)
JOURNAL    21570554
MEDLINE
COMMENT    Contact: Yoshikazu Sugimoto
            Animal Genetics Division
            Shrikraka Institute of Animal Genetics
            Odakura, Nishigo, Nishi-shirakawa, Fukushima 961-8061, Japan
            Tel: 81-248-25-5641
            Fax: 81-248-25-5725
            Email: kazusugi@cocoa.ocn.ne.jp
            This clone was obtained from a polyA-deleted cDNA library.
FEATURES
             source
               1..514
                /organism="Bos taurus"
                /db_xref="taxon:9913"
                /clone="E1C047F06"
                /clone_1lb="Bos taurus cartilage fetus"
                /tissue_type="cartilage"
                /dev_stage="fetus"
                /lab_host="DH10B"
                /note="Vector: pZUL1, Site 1: SalI, Site 2: NotI, poly A
              was deleted from a NotI site"
BASE COUNT 162 a          94 c          184 g          73 t          1 others
ORIGIN

Alignment Scores:
Score:      165          Length:      514
Pred. No.:   49.00       Matches:      8
Percent Similarity: 90.91% Conservative: 2
Best Local Similarity: 72.73% Mismatches:  1
Query Match:  54.44%     Indels:    0
DB:           10         Gaps:      0

US-10-020-441-3 (1-16) x AV600039 (1-514)

Oy      5  ARGASPASnaPrgAsppProSerARgLyAs 15
Db      286 CGAGACGAGAACGACAGGAGCCCGAGAGAGAG 318

```

```

RESULT 12
LOCUS      BJI17850                      549 bp    mRNA       linear   EST-23-JAN-2002
DEFINITION unpublised oligo-capped cdna library, C.elegans L1 stage
ACCESSION  BJI127850
VERSION     BJI127850
KEYWORDS    GI:18288007
SOURCE      Caenorhabditis elegans.
ORGANISM    Caenorhabditis elegans
AUTHORS      Eukaryote; Metazoa; Nematoda; Chromodorea; Rhabdilitia; Rhabdicoidae ;
              Rhabditidae; Peloderinae; Caenorbabdtlis.
REFERENCE   1 (bases 1 to 549)
            Kohara,Y., Shin-i,T., Thierry-Mieg,J., Thierry-Mieg,D., Suzuki,Y.
            and Sugano,S.
            A complementary view of the C.elegans genome
TITLE        Unpublished (2002)
JOURNAL      Contact : Tadasu Shin-i
COMMENT      Center For Genetic Resource Information
             National Institute of Genetics
             111 Yata, Mishima, Shizuoka 411-8540, Japan
             Tel.: 81-559-81-6856
             Fax: 81-559-81-6855
             Email: tshini@gens.nig.ac.jp.
FEATURES
source
         1..549
           /organism="Caenorhabditis elegans"
           /strain="N2"
           /db_xref="taxon:6239"
           /clone="yk1351e07"
           /clone_1kb="unpublished oligo-capped cdna library, C.
               elegans L1 stage"
           /sex="hermaphrodite"
           /tissue_type="whole animal"
           /dev_stage="L1"
           /note="The AD-wmcDNA library was generated with poly(A)+
               RNA isolated from both hermaphrodite and male NZ worms of
               all larval stages, embryos, adults and dauvers and the
               subsequent generation of cDNAs by poly(A) priming. The
               cDNAs were cloned into pPC96"
BASE COUNT  143 a          123 c          153 g          130 t
ORIGIN
Alignment Scores:
Pred. No.:                179                        Length:                  549
Score:                    49.00                       Matches:                          9
Percent Similarity:       71.43%                     Conservative:                      1
Best Local Similarity:    64.29%                     Mismatchches:                     4
Query Match:              54.44%                     Indels:                           0
DB:                         13                            Gaps:                             0

US-10-020-441-3 (1-16) x BJI127850 (1-549)
QY      2  G|YTYYG|YAARGASPA$PASPAP$AARGA$PPROSGRAGLYLS 15
           |||||
Db      169 GGTCGTGGAGAGATGATTGCATCGTGTGATGAGCCGAGACCT 210
           |||||||
RESULT 13
LOCUS      AV606897                      556 bp    mRNA       linear   EST-28-NOV-2001
DEFINITION unpublised oligo-capped cdna clone FIKI037E11
ACCESSION  AV606897
VERSION     AV606897
KEYWORDS    5' , mRNA sequence .
ACCESSION  AV606897.1  GI:9737270
EST.
SOURCE      cow.
ORGANISM    Bos taurus
AUTHORS      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
              Mammalia; Euarchontes; Cetartiodactyla; Ruminantia; Pecora; Bovidea;
              Bovidae; Bovinae; Bos.
REFERENCE   1 (bases 1 to 556)
            Takasuga,A., Hirotsune,S., Itch,R., Jitchazono,A., Suzuki,H., Ago,H.
```

TITLE	and Sugimoto, Y.		
JOURNAL	Establishment of a high throughput EST sequencing system using		
MEDLINE	poly(A) tail-removed cDNA libraries and determination of 36,000		
COMMENT	bovine ESTs		
	Nucleic Acids Res. 29 (22), E108 (2001)		
	21570554		
	Contact: Yoshikazu Sugimoto		
	Animal Genetics Division		
	Shirakawa Institute of Animal Genetics		
	Odkura, Nishigo, Nishi-Shirakawa, Fukushima 961-8061, Japan		
	Tel: 81-248-25-5641		
	Fax: 81-248-25-5725		
	Email: kazusugi@cocoa.ocn.ne.jp		
	Single pass sequencing.		
	This clone was obtained from a polyA-deleted cDNA library.		
FEATURES	Location/Qualifiers		
source	1..556		
	/organism="Bos taurus"		
	/db_xref="taxon:9913"		
	/clone="EIKI037E11"		
	/clone_lib="Bos taurus kidney fetus"		
	/tissue_type="kidney"		
	/dev_stage="fetus"		
	/lab_host="DH10B"		
	/note="Vector: pZL1, Site_1: SalI, Site_2: NotI, Poly A		
	was deleted from a NotI site"		
BASE COUNT	177 a 104 c 199 g 76 t		
ORIGIN			
Alignment Scores:			
Pred. No.:	182	Length:	556
Score:	49.00	Matches:	8
Percent Similarity:	90.91%	Conservative:	2
Best Local Similarity:	72.73%	Mismatches:	1
Query Match:	54.44%	Indels:	0
DB:	10	Gaps:	0
US-10-020-441-3 (1-16) x AV606897 (1-556)			
QY	5 ATGAAPASpASnspARgAspProSeArGtLyS 15		
Db	199 CGAGCGAGAACGACGAGGCCAGAGAGAGAG 231		
LOCUS	BB642648		
DEFINITION	BB642648 RIKEN full-length enriched, adult retina Mus musculus cDNA		
VERSION	BB642648		
KEYWORDS	BB642648.1 GI:15401968		
ORGANISM	EST.		
REFERENCE	house mouse.		
AUTHORS	Mus musculus		
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.		
	1 (bases 1 to 668)		
	Arakawa, T., Carninci, P., Fukuda, S., Furuno, M., Hanagaki, T., Hara, A.,		
	Hiramoto, K., Hori, F., Ishii, Y., Ito, M., Kawai, J., Kono, H., Kouda		
	M., Koya, S., Matsuyama, T., Miyazaki, A., Nomura, K., Ohno, M.,		
	Okazaki, Y., Okido, T., Saito, R., Sakai, C., Sakai, K., Sano, H., Sasaki		
	D., Shibata, K., Shinagawa, A., Shiraki, T., Sogabe, Y., Suzuki, H.,		
	Tagami, M., Tagawa, A., Takehashi, F., Takeda, Y., Tanaka, T., Toyata, T.,		
	Muramatsu, M., and Hayashizaki, Y.		
	RIKEN Mouse ESTs (Arakawa, T., et al. 2001)		
	Unpublished (2001)		
TITLE	Contact: Yoshihide Hayashizaki		
JOURNAL	Laboratory for Genome Exploration Research Group, RIKEN Genomic		
COMMENT	Sciences Center (GSC), Yokohama Institute		
	The Institute of Physical and Chemical Research (RIKEN)		
	1-7-22 Saitama-cho, Tsukuba-shi, Ibaraki 305-3858, Japan		
	Tel: 81-45-503-9222		
	Fax: 81-45-503-9216		
	Email: genome-res@gsc.riken.go.jp,		

URL:<http://genome.gsc.riken.go.jp/>  
Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K., Itoh,M., Kono,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.  
Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes. Genome Res. 10 (10), 1617-1630 (2000)  
wagi,K., Fujiwaka,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E., Wataniki,M., Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsunura,S., Kawai,U., Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and Hayashizaki,Y.  
RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multicapillary sequencer. Genome Res. 10 (11), 1757-1771 (2000)  
Kono,H., Fukunishi,Y., Shibata,K., Itoh,M., Carninci,P., Sugahara,Y. and Hayashizaki,Y.  
Computer-based methods for the mouse full-length cDNA encyclopedia: real-time sequence clustering for construction of a nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)  
Yamanaka,I., Kiyosawa,H., Kondo,S., Saito,T., Shingawa,A., Aizawa,K., Fukuda,S., Hara,A., Iton,M., Kawai,U., Shibata,K., Atakawa,T., Ishii,Y. and Hayashizaki,Y.  
Mapping of 19032 mouse cDNAs on mouse chromosomes. J. Struct. Func. Genomics 2 pre, L72-L86 (2001)  
Please visit our web site (<http://genome.gsc.riken.go.jp>) for further details.

### FEATURES

source Location/Qualifiers  
1..668 /organism="Mus musculus"  
/db\_xref="taxon:10090"  
/clone\_id="RIKEN full-length enriched, adult retina"  
/tissue\_type="retina"  
/dev\_stage="adult"  
/lab\_host="DH10B"  
/note="Site 1: Sali; Site 2: BamHI; cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. 1st strand cDNA was primed with a primer [5'  
GAGAGAGAGATCCACAGCTCTTTTTCCTTTTTTTTNN 3'] , cDNA was prepared by using trehalose thermo-activated reverse transcriptase and subsequently enriched for full-length by cap-trapper. cDNA went through one round of normalization to Rot = 20.0 and subtraction to Rot = 459.0. Second strand cDNA was prepared with the primer adapter of sequence [5' GAGAGAGAGATTCTCGACTTAATTAATTAATCCCCCCCCCCC 3']. cDNA was cleaved with XhoI and BamHI. Vector: a modified pBluescript KS(+) after bulk excision from Lambda FLC I. ~Retina RNA was provided by Stefano Gustinich, Department of Neurobiology, Harvard Medical School, 220 Longwood Ave., Boston, MA02115, USA, whose assistance we gratefully acknowledge."

BASE COUNT	138 a	191 c	221 g	117 t	1 others
ORIGIN					

### Alignment Scores:

Pred. No.:	229	Length:	668
Score:	49.00	Matches:	9
Percent Similarity:	76.92%	Conservative:	1
Best Local Similarity:	69.23%	Mismatches:	3
Query Match:	54.44%	Indels:	0
DB:	10	Gaps:	0

US-10-020-441-3 (1-16) x BB642648 (1-668)

```

Oy      4 G1YAAGAaPaenAaSPaGPProSeArqylSaSn 16
          |||||:::|||||:::|||||:::|||||
Db     38 GGAAGAGACGATACGATCCAGAAAGCTCTCAAGACCAAC 76
    
```

RESULT 15

LOCUS	BB642873	682 bp	mRNA	linear	EST 26-OCT-2001
DEFINITION	BB642873	RIKEN full-length enriched, adult retina	Mus musculus cDNA		
ACCESSION	BB642873	clone A930104M11 5', mRNA sequence.			
VERSION	BB642873				
KEYWORDS	BB642873.1	GI:16477631			
SOURCE	EST.				
ORGANISM	house mouse.				
REFERENCE	Mus musculus				
AUTHORS	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.				
	1 (bases 1 to 682)				
	Arakawa,T., Carninci,P., Fukuda,S., Furuno,M., Hanagaki,T., Hara,A., Hiramoto,K., Hori,F., Ishii,Y., Ito,M., Kawai,U., Komoto,H., Kouda,M., Koya,S., Matsuyama,T., Miyazaki,A., Nomura,K., Ohno,M., Okazaki,Y., Okido,T., Saito,R., Sakai,C., Sakai,K., Sano,H., Sasaki,D., Shibata,K., Shingawa,A., Shiraki,T., Sogabe,Y., Suzuki,H., Tagami,M., Tagawa,A., Takahashi,F., Takeda,Y., Tanaka,T., Toya,T., Muramatsu,M. and Hayashizaki,Y.				
TITLE	RIKEN Mouse ESTs (Arakawa,T., et al. 2001)				
JOURNAL	Unpublished (2001)				
COMMENT	Contact: Yoshihide Hayashizaki Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center(GSC), Yokohama Institute The Institute of Physical and Chemical Research (RIKEN) 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan Tel: 81-45-503-9222 Fax: 81-45-503-9216 Email: genome-res@gsr.riken.go.jp, url: http://genome.gsc.riken.go.jp/ Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K., Itoh,M., Komoto,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y. Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes. Genome Res. 10 (10), 1617-1630 (2000) Wagii,K., Fujiwara,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E., Wakahiki,M., Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsuura,S., Kawai,U., Okazaki,Y., Muramatsu,M., Inoue,Y., Kita,A. and Hayashizaki,Y. RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multicapillary sequencer. Genome Res. 10 (11), 1757-1771 (2000) Komoto,H., Fukunishi,Y., Shibata,K., Itoh,M., Carninci,P., Sugahara,Y. and Hayashizaki,Y. Computer-based methods for the mouse full-length cDNA encyclopedia: real-time sequence clustering for construction of a nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001) Kondo,S., Shingawa,A., Saito,T., Kiyosawa,H., Yamanaka,I., Aizawa,K., Fukuda,S., Hara,A., Itoh,M., Kawai,U., Shibata,K. and Hayashizaki,Y. Computational Analysis of Full-length Mouse cDNAs Compared with Human Genome Sequences. Mamm. Genome. 12, 673-677 (2001) Please visit our web site ( <a href="http://genome.gsc.riken.go.jp">http://genome.gsc.riken.go.jp</a> ) for further details. e mouse tissues. Location/Qualifiers 1. 682 /organism="Mus musculus" /db_xref="taxon:10090" /clone="A930104M11" /clone_lib="RIKEN full-length enriched, adult retina" /tissue_type="retina" /dev_stage="adult" /lab_host="DH10B" /note="Site 1: Sali; Site 2: BamHI. cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. 1st strand cDNA was primed with a primer [5', GAGAGAGGAGAGATCCAGAGCTCTTTTCTTTTCTTTT 3'], cDNA was prepared by using triethylalcohol thermo-activated reverse				

transcriptase and subsequently enriched for full-length by cap-trapper. cDNA went through one round of normalization to Rot = 20.0 and subtraction to Rot = 459.0. Second strand cDNA was prepared with the primer adapter of sequence (5'GAGAGAGATTCTCGACTTAAATTAAATCCCCCCCCC 3'). cDNA was cleaved with XhoI and BamHI. Vector: a modified pBluescript KS(+) after bulk excision from Lambda PLC I. ~Retina RNA was provided by Stefano Gustinich, Department of Neurobiology, Harvard Medical School, 220 Longwood Ave., Boston, MA02115, USA, whose assistance we gratefully acknowledge."	
BASE COUNT	142 a 192 c 223 g 120 t 5 others
ORIGIN	
Alignment Scores:	
Pred. No.:	235
Score:	49.00
Percent Similarity:	76.92%
Best Local Similarity:	69.23%
Query Match:	54.44%
DB:	10 Gaps: 0
US-10-020-441-3 (1-16) x BB642873 (1-682)	
CY	4 G1YARGASPAASPARCASPProSerArgLysAsn 16
Db	38 GGAAGAGACGATACGATCGAGAGCCTCAAGAGCAAC 76
RESULT 16	
BH839335	
LOCUS	737 bp DNA linear GSS 28-MAY-2002
DEFINITION	LMCRI50011D08f Zea mays L. Zea mays genomic clone LMCRI50011D08f,
ACCESSION	BH839335
VERSION	BH839335.1 GI:21237350
KEYWORDS	GSS.
SOURCE	Zea mays.
ORGANISM	Zea mays.
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC clade; Panicoideae; Andropogoneae; Zea.	
REFERENCE	1 (bases 1 to 737)
AUTHORS	Kim,S.W., Yu,Y., Lee,M.C., Main,D. and Wing,R.A.
TITLE	Methyl-filtration genomic sequence from maize
JOURNAL	Unpublished (2002)
COMMENT	Contact: Wing RA Clemson University Genomics Institute Clemson University 100 Jordan Hall, Clemson, SC 29634, USA Tel: 864 656 7288 Fax: 864 656 4293 Email: rwing@clemson.edu Total High Quality bases = 509 Seq primer: TAATACGACTCCTATATAGGC Class: shotgun High quality sequence start: 25 High quality sequence stop: 682.
FEATURES	Location/Qualifiers
source	1..737 /organism="Zea mays" /strain="B73" /db_xref="taxon:4577" /clone="LMCRI50011D08f" /clone_1fb="Zea mays L." /tissue_type="Leaf" /lab_host="DH10B" /note="Vector: pGEM-T easy, Site 1: Mcr BC; Methyl-filtration library, Nuclei DNA was completely digested with Mcr BC, size fractionated and transformed to E.Coli DH10B."
BASE COUNT	157 a 253 c 214 g 111 t 2 others
ORIGIN	



REFERENCE 2 (bases 1 to 372)  
AUTHORS Roest-Crolius,H., Jalllon,O., Dasilva,C., Fikames,C., Fisher,C.,  
Bonneau,I., Billault,A., Quetier,F., Saurin,W., Bernot,A. and  
Weissenbach,J.  
TITLE Characterization and repeat analysis of the compact genome of the  
freshwater pufferfish Tetraodon nigroviridis  
JOURNAL Unpublished  
REFERENCE 3 (bases 1 to 372)  
AUTHORS Genoscope.  
TITLE Direct Submission  
JOURNAL Submitted (12-APR-2000)  
COMMENT This sequence is a single read and was generated as part of a large  
scale clone-end sequencing project of the Tetraodon nigroviridis  
genome. For more information, please take a look at  
http://www.genoscope.cns.fr/Tetraodon.  
FEATURES  
SOURCE location/Qualifiers  
1..372  
/organism="Tetraodon nigroviridis"  
/db\_xref="taxon:99883"  
/clone="061G11"  
/clone\_lib="G"  
/note="Genoscope sequence ID : COB0061AD06L1-end : T7"  
BASE COUNT 75 a 104 c 124 g 67 t 2 others  
ORIGIN  
Alignment Scores:  
Pred. No.: 164 Length: 372  
Score: 48.00 Matches: 8  
Percent Similarity: 73.33% Conservative: 3  
Best Local Similarity: 53.33% Mismatches: 4  
Query Match: 53.33% Indels: 0  
DB: 17 Gaps: 0  
US-10-020-441-3 (1-16) x CNS030Z9 (1-372)  
Qy 1 ValGIYTYRGlyARgAspAsnAspARgAspProSerArgLys 15  
Db 141 ATGTCTATGGAGGAGGACGATCGACTCTGACGACGCCGACAG 185  
RESULT 20  
AL376173/c 461 bp mRNA linear EST 03-AUG-2000  
LOCUS MCB21G08P1 MCBB Medicago truncatula cDNA clone MCB21G08 T3, mRNA  
DEFINITION  
ACCESSION AL376173  
VERSION AL376173  
KEYWORDS GI:9675925  
SOURCE EST.  
ORGANISM barrel medic.  
Medicago truncatula  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Trifoliaceae;  
Medicago.  
REFERENCE 1 (bases 1 to 461)  
AUTHORS Journef,E.P., Crespeau,H., van-Tuinen,D., Gouzy,J., Jalllon,O.,  
Niebel,A., Carreau,V., Chataigner,O., Kahn,D., Gianinazzi-Pearson  
, V. and Gamas,P.  
TITLE Medicago truncatula ESTs from Sinorhizobium meliloti-induced root  
nodules  
JOURNAL Unpublished (2000)  
COMMENT Genoscope - Centre National de Sequencage  
BP 191 91006 EVRY cedex - France  
Email: segref@genoscope.cns.fr, Web : www.genoscope.cns.fr  
Contact : Pascal Gamas and Etienne-Pascal Journef, Laboratoire de  
Biologie Moleculaire des Relations Plantes-Microorganismes,  
CNRS-INRA, BP 27 31326 Castanet-Tolosan Cedex, France (Email :  
Mt-est@toulouse.inra.fr Website :  
http://sequence.toulouse.inra.fr/Mtruncatula.html).  
FEATURES  
SOURCE location/Qualifiers  
1..461  
/organism="Medicago truncatula"  
/cultivar="Jemalong"  
/clone="MCB44C05"

/db\_xref="taxon:3880"  
/clone="MCB21G08"  
/clone\_lib="MCBB"  
/tissue\_type="symbiotic root nodules"  
/dev\_stage="harvested 4 days post inoculation with  
Sinorhizobium meliloti"  
/note="Vector: pBluescript pSK, Site\_1: EcoRI, Site\_2:  
XhoI; Plants were grown in an aeroponic chamber on  
nitrogen-rich medium for 21 days. Three days before  
inoculation with Sinorhizobium meliloti, the medium was  
replaced by N-free medium. Root nodules (+ short adjacent  
root segments) were harvested 4 days post inoculation.  
cDNA was prepared from polyA+ enriched RNA. The cDNA was  
directionally ligated into Uni-zap XR vector from  
Stratagene and packaged using Gigapack Gold packaging  
extracts. Plasmids containing cDNA inserts were  
mass-excised from phage stocks using ExsacII helper phage  
and propagated in SOLR cells. Clone ordering and  
sequencing was performed by the Centre National de  
Sequencage (Genoscope, Evry, France)."  
BASE COUNT 137 a 87 c 109 g 128 t  
ORIGIN  
Alignment Scores:  
Pred. No.: 213 Length: 461  
Score: 48.00 Matches: 8  
Percent Similarity: 83.33% Conservative: 2  
Best Local Similarity: 66.67% Mismatches: 2  
Query Match: 53.33% Indels: 0  
DB: 9 Gaps: 0  
US-10-020-441-3 (1-16) x AL376173 (1-461)  
Qy 5 ARgAspAsnAspARgAspProSerArgLysAsn 16  
Db 194 CGTGACGATCAACCCAGAGATCTTCCAAACGAAT 159  
RESULT 21  
AL379254/c 477 bp mRNA linear EST 03-AUG-2000  
LOCUS MCB44C05P1 MCBB Medicago truncatula cDNA clone MCB44C05 T3, mRNA  
DEFINITION  
ACCESSION AL379254  
VERSION AL379254  
KEYWORDS GI:9679006  
SOURCE EST.  
ORGANISM barrel medic.  
Medicago truncatula  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Trifoliaceae;  
Medicago.  
REFERENCE 1 (bases 1 to 477)  
AUTHORS Journef,E.P., Crespeau,H., van-Tuinen,D., Gouzy,J., Jalllon,O.,  
Niebel,A., Carreau,V., Chataigner,O., Kahn,D., Gianinazzi-Pearson  
, V. and Gamas,P.  
TITLE Medicago truncatula ESTs from Sinorhizobium meliloti-induced root  
nodules  
JOURNAL Unpublished (2000)  
COMMENT Genoscope - Centre National de Sequencage  
BP 191 91006 EVRY cedex - France  
Email: segref@genoscope.cns.fr, Web : www.genoscope.cns.fr  
Contact : Pascal Gamas and Etienne-Pascal Journef, Laboratoire de  
Biologie Moleculaire des Relations Plantes-Microorganismes,  
CNRS-INRA, BP 27 31326 Castanet-Tolosan Cedex, France (Email :  
Mt-est@toulouse.inra.fr Website :  
http://sequence.toulouse.inra.fr/Mtruncatula.html).  
FEATURES  
SOURCE location/Qualifiers  
1..477  
/organism="Medicago truncatula"  
/cultivar="Jemalong"  
/db\_xref="taxon:3880"  
/clone="MCB44C05"

```

/clone_lib="MEBR"
/tissue_type="symbiotic root nodules"
/dev_stage="harvested 4 days post inoculation with
Sinorhizobium meliloti"
/notes="vector: pBluescript pSK; Site 1: EcoRI; Site 2:
XhoI; Plants were grown in an aeroponic chamber on
nitrogen-rich medium for 21 days. Three days before
inoculation with Sinorhizobium meliloti, the medium was
replaced by N-free medium. Root nodules (+ short adjacent
root segments) were harvested 4 days post inoculation.
cDNA was prepared from polyA+ enriched RNA. The cDNA was
directionally ligated into Uni-zap XR vector from
Stratagene and packaged using Gigapack Gold packaging
extracts. Plasmids containing cDNA inserts were
mass-excised from phage stocks using Exsist helper phage
and propagated in SGR cells. Clone ordering and
sequencing was performed by the Centre National de
Sequencage (Genoscope, Evry, France)."
```

BASE COUNT 136 a 98 c 92 g 151 t

ORIGIN

Alignment Scores:

Pred. NO.:	223	Length:	477
Score:	48.00	Matches:	8
Percent Similarity:	83.33%	Conservative:	2
Best Local Similarity:	66.67%	Mismatches:	2
Query Match:	53.33%	Indels:	0
DB:	9	Gaps:	0

US-10-020-441-3 (1-16) x AL379254 (1-477)

Oy 5 ArgadpApAnspArqApProSeArglyAaen 16  
 |||||  
 Db 132 CGTGCAGATCAACCCAGATCTCTCCAAACCAAT 97  
 |||||

RESULT 22

LOCUS	AO964575	500 bp	DNA	linear	GSS 28-JAN-2000
DEFINITION	LERGX17TR LERG Arabidopsis thaliana genomic clone LERGX17, DNA sequence.				
ACCESSION	AO964575				
VERSION	AO964575.1	GI:6792276			
KEYWORDS	GSS.				
SOURCE	thale cress.				
ORGANISM	Arabidopsis thaliana				
REFERENCE	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.				
AUTHORS	1 (bases 1 to 500)				
REFERENCE	Buell, C.R., Lin, X., Pat, G., Barnstead, M., Bowman, C., Uterbach, T., Feldygun, T., Liang, F., Creasy, T. and Fraser, C.M.				
COMMENT	Genomic survey sequencing of Landsberg erecta ecotype of Arabidopsis thaliana and identification of sequence-based polymorphisms				
JOURNAL	Unpublished (2000)				
CONTACT	Contact: Xiaoying Lin				
INSTITUTE	The Institute for Genomic Research				
ADDRESS	9712 Medical Center Dr., Rockville, MD 20850, USA				
TELEPHONE	Tel: 301 838 0200				
FAX	Fax: 301 838 0208				
EMAIL	Email: ael@tigr.org				
FOR ADDITIONAL INFORMATION	See <a href="http://www.tigr.org/tdb/at/at.html">http://www.tigr.org/tdb/at/at.html</a>				
SEG PRIMER	TR				
CLASS	shotgun.				

FEATURES

Source

Location/Qualifiers

1..500

/organism="Arabidopsis thaliana"

/strain="Landsberg erecta"

/db\_xref="taxon:3702"

/clone="LERGX17"

/clone\_lib="LERG"

/note="Organ: Leaf; Vector: pUC19K; Total genomic DNA was sheared to 0.4-0.7 Kbp before ligation."

ORIGIN	BASE COUNT	190 a	77 c	96 g	137 t
Alignment Scores:					
Pred. No.:	236	Length: 500			
Score:	48.00	Matches: 8			
Percent Similarity:	84.62%	Conservative: 3			
Best Local Similarity:	61.54%	Mismatch: 2			
Query Match:	53.33%	Indels: 0			
DB:	17	Gaps: 0			
US-10-020-441-3 (1-16) x AQ964575 (1-500)					
QY	2	G1YTYG1YARGAsPAsPAsnAsPARGAsPProSeArG 14			
DB	255	GGATTGGAGGTAAAGATCAAGATCAAGATCAATCAATAA 293			
RESULT 23					
LOCUS	AQ964574/c	521 bp	DNA	linear	GSS 28-JAN-2000
DEFINITION	LERGX17TF LERG Arabidopsis thaliana genomic clone LERGX17, DNA sequence.				
ACCESSION	AQ964574				
VERSION	AQ964574.1	GI:6792275			
KEYWORDS	GSS.				
SOURCE	thale cress.				
ORGANISM	Arabidopsis thaliana				
REFERENCE	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.				
AUTHORS	Buell, C.R., Lin, X., Pai, G., Barnstead, M., Bowman, C., Uterbach, T., Feldblum, T., Liang, F., Creasy, T. and Frazer, C.M.				
TITLE	Genomic survey sequencing of Landsberg erecta ecotype of Arabidopsis thaliana and identification of sequence-based polymorphisms				
JOURNAL	Unpublished (2000)				
COMMENT	Contact: Xiaoying Lin The Institute for Genomic Research 9712 Medical Center Dr., Rockville, MD 20850, USA Tel: 301 838 0200 Fax: 301 838 0208 Email: atc@tigr.org For additional information, see <a href="http://www.tigr.org/cdb/at/at.html">http://www.tigr.org/cdb/at/at.html</a> Seq primer: TF Class: shotgun.				
FEATURES	location/Qualifiers				
SOURCE	1..521				
	/organism="Arabidopsis thaliana"				
	/strain="Landsberg erecta"				
	/db_xref="taxon:3702"				
	/clone="LERGX17"				
	/clone_id="LERG"				
	/note="Organ: Leaf; Vector: pUC19UK; Total genomic DNA was sheared to 0.4-0.7 kbp before ligation."				
BASE COUNT	146 a	97 c	79 g	199 t	
ORIGIN					
Alignment Scores:					
Pred. No.:	248	Length: 521			
Score:	48.00	Matches: 8			
Percent Similarity:	84.62%	Conservative: 3			
Best Local Similarity:	61.54%	Mismatch: 2			
Query Match:	53.33%	Indels: 0			
DB:	17	Gaps: 0			
US-10-020-441-3 (1-16) x AQ964574 (1-521)					
QY	2	G1YTYG1YARGAsPAsPAsnAsPARGAsPProSeArG 14			
DB	319	GGATTGGAGGTAAAGATCAAGATCAAGATCAATCAATAA 281			
RESULT 24					

AL379255/c 525 bp mRNA linear EST 03-AUG-2000  
LOCUS MCBB4C05R1 MtBB Medicago truncatula cDNA clone MCBB4C05 T7, mRNA  
DEFINITION sequence.  
ACCESSION AL379255  
VERSION AL379255.1 GI:9679007  
KEYWORDS EST.  
SOURCE barrel medic.  
ORGANISM Medicago truncatula  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
Rosidae; eustosids I; Fabales; Fabaceae; Papilionoideae; Trifoliaceae;  
Medicago.  
REFERENCE 1 (bases 1 to 525)  
AUTHORS Journal,E.P., Crespeau,H., van-Tuinen,D., Gouzy,U., Jaillon,O.,  
Niebel,A., Carreau,V., Chatagnier,O., Kahn,D., Gianinazzi-Pearson  
,V. and Gamas,P.  
TITLE Medicago truncatula ESTs from Sinorhizobium meliloti-induced root  
nodules  
JOURNAL Unpublished (2000)  
COMMENT Contact: Genoscope  
Genoscope - Centre National de Sequencage  
BP 191 91006 Evry cedex - France  
Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr  
Contact : Pascal Gamas and Etienne-Pascal Journet, Laboratoire de  
Biologie Moleculaire des Relations Plantes-Microorganismes,  
CNRS-INRA, BP 27 31326 Castanet-Tolosan Cedex, France (Email :  
Mc-est@toulouse.inra.fr Website :  
http://sequence.toulouse.inra.fr/Mtruncatula.html).  
Location/Qualifiers  
1..525  
/organism="Medicago truncatula"  
/cultivar="Jemalong"  
/db\_xref="taxon:3880"  
/clone="MCBB4C05"  
/clone\_1fb="MCBB"  
/tissue\_type="symbiotic root nodules"  
/dev\_stage="harvested 4 days post inoculation with  
Sinorhizobium meliloti"  
/note="Vector: pBluescript PSK; Site\_1: EcoRI; Site\_2:  
XhoI; Plants were grown in an aeroponic chamber on  
nitrogen-rich medium for 21 days. Three days before  
inoculation with Sinorhizobium meliloti, the medium was  
replaced by N-free medium. Root nodules (+ short adjacent  
root segments) were harvested 4 days post inoculation.  
cDNA was prepared from polyA+ enriched RNA. The cDNA was  
directionally ligated into Uni-zap XR vector from  
Stratagene and packaged using Gigapack Gold packaging  
extracts. Plasmids containing cDNA inserts were  
mass-excised from phage stocks using Exasit helper phage  
and propagated in SOLR cells. Clone ordering and  
sequencing was performed by the Centre National de  
Sequencage (Genoscope, Evry, France)."  
BASE COUNT 163 a 104 c 94 g 164 t  
ORIGIN  
Alignment Scores:  
Pred. No.: 251 Length: 525  
Score: 48.00 Matches: 8  
Percent Similarity: 83.33% Conservative: 2  
Best Local Similarity: 66.67% Mismatches: 2  
Query Match: 53.33% Indels: 0  
DB: 9 Gaps: 0  
US-10-020-441-3 (1-16) x AL379255 (1-525)  
Qy 5 ArgaspaspaasnspargspProserArglysaasn 16  
|||||  
Db 114 CGTGACGATCAACCGACGATCTTCCAAACGAAT 79  
|||||  
RESULT 25  
LOCUS B1920718 560 bp mRNA linear EST 17-OCT-2001

EST540653 potato microtubers, in vitro-grown Solanum tuberosum cDNA  
clone cSTE26E16 5' end, mRNA sequence.  
ACCESSION B1920718  
VERSION B1920718.1 GI:16216746  
KEYWORDS EST.  
SOURCE potato.  
ORGANISM Solanum tuberosum  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
Asteridae; euasterids I; Solanales; Solanales; Solanales; Solanum.  
REFERENCE 1 (bases 1 to 560)  
AUTHORS van der Hoeven,R., Bezzerides,J., Bachem,C., Visser,R., Karamycheva  
,S.A., Teal,J., Van Aken,S., Uteerback,T., Chieningo,A., Bougri,O.,  
Buell,C.R., Roming,C., Tanksley,S. and Baker,B.  
TITLE Generation of ESTs from in vitro grown microtubers (2001b)  
JOURNAL Unpublished (2001)  
COMMENT Contact: Research Genetics, Libraries Division  
Tel: 1-800-711-6195  
Email: cdna@resgen.com  
For clone info: please contact Research Genetics, Libraries  
Division tel 1-800-711-6195, email cdna@resgen.com  
Seq primer: T3.  
Location/Qualifiers  
1..560  
/organism="Solanum tuberosum"  
/cultivar="Bintje"  
/db\_xref="taxon:4113"  
/clone="cSTE26E16"  
/clone\_1fb="potato microtubers, in vitro-grown"  
/tissue\_type="axillary buds of stem explants; growing  
sink-tubers"  
/dev\_stage="7, 8 and 10 days"  
/lab\_host="SOLR"  
/note="Vector: pBluescript SK(-); Site\_1: EcoRI; Site\_2:  
XhoI; Tissue supplied by Christian Bachem and Richard  
Visser (Department of Plant Breeding, Wageningen  
University, The Netherlands); sequencing by The Institute  
for Genomic Research. The cDNA libraries will attempt to  
capture the induction and initiation/initial growth of the  
tuber in an in vitro system as described in Bachem et al.  
(Plant Journal, 1996). Small microtubers develop from  
axillary buds attached to stem explants when placed on a  
high sucrose medium (10%). Visible morphological changes  
occur synchronously at day five in the axillary buds. The  
first library, cSTA (1-20) consists of axillary buds  
harvested on days 1-3. This targets those genes involved  
in induction of the microtubers. The following libraries,  
cSTA (21-40) and cSTA (41-60), capture genes involved in  
tuber initiation and outgrowth. This library is noted as  
P3 in Tanksley lab notebooks."  
BASE COUNT 184 a 106 c 146 g 124 t  
ORIGIN  
Alignment Scores:  
Pred. No.: 271 Length: 560  
Score: 48.00 Matches: 8  
Percent Similarity: 64.29% Conservative: 1  
Best Local Similarity: 57.14% Mismatches: 5  
Query Match: 53.33% Indels: 0  
DB: 13 Gaps: 0  
US-10-020-441-3 (1-16) x B1920718 (1-560)  
Qy 2 GLYTYTGTGARGASPASPASPARGSPProserArglyys 15  
|||||  
Db 40 GGACATGGCAACGACGACGCGCGACCCAGAAAGAGAAA 81  
|||||  
RESULT 26  
LOCUS AM679964 573 bp mRNA linear EST 19-JUN-2000  
DEFINITION WS1\_34\_C05\_g1\_A002 Water-stressed 1 (WS1) Sorghum bicolor cDNA,  
mRNA sequence.  
ACCESSION AM679964



VERSION AM679964.1 GI:7553741  
 KEYWORDS EST.  
 SOURCE Sorghum.  
 ORGANISM *Sorghum bicolor*  
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 Spermatophytes; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC  
 clade; Panicoideae; Andropogoneae; Sorghum.  
 REFERENCE 1 (bases 1 to 573)  
 AUTHORS Cordonnier-Pratt,M.-M., Gingle,A., Marsala,C., Sudman,M. and Pratt  
 ,L.H.  
 TITLE An EST database from Sorghum: water-stressed plants  
 JOURNAL Unpublished (2000)  
 COMMENT Contact: Cordonnier-Pratt MM  
 Laboratory for Genomics and Bioinformatics  
 The University of Georgia, Department of Plant Biology  
 Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA  
 Tel: 706 542 1860  
 Fax: 706 583 0210  
 Email: mmprratt@ga.edu  
 Sequences have been trimmed to exclude PolyA, vector and regions  
 below Phred quality 16. The threshold for highest quality sequence  
 is 20.  
 Seq primer: T7  
 High quality sequence start: 42  
 High quality sequence stop: 573  
 PolyA=yes.  
 FEATURES  
 source Location/Qualifiers  
 1..573  
 /organism="Sorghum bicolor"  
 /db\_xref="taxon:4558"  
 /clone\_lib="Water-stressed 1 (WS1)"  
 /note="Organ: Mix of 5-week old plants on days 7 & 8 after  
 water was withheld; Vector: Lambda Zap; Site 1: XhoI;  
 Site 2: EcoRI; The library was made from poly-A RNA in the  
 cloning vector lambda Zap II. Clones to be sequenced were  
 prepared by mass excision."  
 BASE COUNT 150 a 106 c 157 g 160 t  
 ORIGIN  
 Alignment Scores:  
 Pred. NO.: 279 Length: 573  
 Score: 48.00 Matches: 7  
 Percent Similarity: 81.25% Conservative: 6  
 Best Local Similarity: 43.75% Mismatches: 3  
 Query Match: 53.33% Indels: 0  
 DB: Gaps: 0  
 US-10-020-441-3 (1-16) x AM679964 (1-573)  
 QY 1 VALGIYTYRGIIYARGApApApApApApApAProSeArGLyAsn 16  
 :::::::::::::::::::::::::::::::::::::::  
 Db 228 ATCGGCTTCGGAAGCAGCATGATGATAGGAAACGACGACAAAGGAC 275  
 :::::::::::::::::::::::::::::::::::::::  
 RESULT 27  
 LOCUS BM108979 713 bp mRNA linear EST 26-NOV-2001  
 DEFINITION Solanum tuberosum roots Solanum tuberosum cDNA clone CPRO2J6 5' end,  
 mRNA sequence.  
 ACCESSION BM108979  
 VERSION BM108979.1 GI:17069573  
 KEYWORDS EST.  
 SOURCE  
 ORGANISM Solanum tuberosum  
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 Spermatophytes; Magnoliophyta; eudicotyledons; core eudicots;  
 Asteridae; euasterids I; Solanales; Solanaceae; Solanum.  
 REFERENCE 1 (bases 1 to 713)  
 AUTHORS van der Hoeven,R., Sun,H., Karanycheva,S.A., Tsai,J., Van Aken,S.,  
 Utecherack,T., Chiemingo,A., Bougri,O., Buell,C.R., Romning,C.,  
 Tanksley,S. and Baker,B.  
 TITLE Generation of ESTs from potato roots  
 JOURNAL Unpublished (2001)  
 COMMENT Contact: Research Genetics, Libraries Division

Tel.: 1-800-711-6195  
Email: cdna@resgen.com  
For clone info: please contact Research Genetics Libraries  
Division tel 1-800-711-6195, email cdna@resgen.com  
Seq primer: T3.

FEATURES

SOURCE

Location/Qualifiers

1..713

/organism="Solanium tuberosum"

/cultivar="Kennebec"

/db\_xref="taxon:4113"

/clone="CPRO2J6"

/clone\_1lb="potato roots"

/tissue\_type="roots"

/dev\_stage="in vitro grown stem cuttings"

/lab\_host="SDR"

/note="Vector: pBluescript SK(-); Site 1: EcoRI; Site 2:  
XhoI; supplier: Cornell University, Tanksley lab;  
sequencing: The Institute for Genomic Research. Roots were  
isolated from in vitro grown stem cuttings on CM medium.  
Roots were isolated two weeks after placing the stem  
cuttings from in vitro grown plants on medium."

BASE COUNT

218 a 159 c 146 g 190 t

ORIGIN

Alignment Scores:

Pred. No.: 366 Length: 713

Score: 48.00 Matches: 8

Percent Similarity: 76.92% Conservative: 2

Percent Local Similarity: 61.54% Mismatches: 3

Query Match: 53.33% Indels: 0

DB: 13 Gaps: 0

US-10-020-441-3 (1-16) x BM108979 (1-713)

Qy 1 ValGIYTYrGlyArgaspAspAsnApArgaspprosEr 13  
:::||||| |||||:::|||||

Dd 560 ATGGAGTATGGAGTGATGATGATGATGATGATGCCTCCCGAC 598

RESULT 28

AUI76144/c

LOCUS

DEFINITION AUI76144 919 bp mRNA linear EST 05-JUN-2001

AUI76144 Marsipusenus japonicus adult Marsipusenus japonicus CDNA

Clone PJV446 reverse similar to Tachyples tridentatus coagulation

factor G-beta chain precursor (B479878), mRNA sequence.

AUI76144

AUI76144.1 GI:14622551

EST.

Marsipusenus japonicus

Marsipusenus japonicus

Eukaryota; Metazoa; Arthropoda; Crustacea; Malacostraca;

Eunalaicostreca; Eucarida; Decapoda; Dendrobranchiata; Penaeidea;

Penaeidae; Marsipusenus.

1 (bases 1 to 919)

Rojitnakorn,J., Hirono,I., Aoki,T., Itami,T. and Takahashi,Y

Gene expression in non-infected and viral infected hemocytes of

kuruma prawn (Penaeus japonicus)

Unpublished (2001)

Contact: Takashi Aoki

Aquatic Bioesciences

Tokyo University of Fisheries

Konan 4-5-7, Tokyo, Minato-ku 108-8477, Japan

Tel: 81-3-5463-0689

Fax: 81-3-5463-0690

Email: aoki@okyo-u-fish.ac.jp.

JOURNAL

COMMENT

location/Qualifiers

1..919

/organism="Marsipusenus japonicus"

/db\_xref="taxon:27405"

/clone="PJV446"

/clone\_1lb="Marsipusenus japonicus adult"

/cell\_type="hemocytes"

/dev\_stage="adult"

/note="Vector: lambda ZAP II"

```

BASE COUNT      247 a      214 c      230 g      227 t      1 others
ORIGIN
Alignment Scores:
Pred. No.:      501      Length:      919
Score:          48.00      Matches:      9
Percent Similarity: 71.43%      Conservative: 1
Best Local Similarity: 64.29%      Mismatches: 4
Query Match:     53.33%      Indels:      0
DB:              9      Gaps:      0

US-10-020-441-3 (1-16) x AUI76144 (1-919)

Qy      1  ValGlyTYrGlyArgAspAsnAspArgAspProSerArg 14
Db      572  GTCCGCTACGGTCGTACGTATCCGACGACGCGATCGAAM 531

RESULT 29
LOCUS      AG083636      919 bp      DNA      linear      GSS 03-NOV-2001
DEFINITION Pan troglodytes DNA, clone: PTB-081C06.R, genomic survey sequence.
ACCESSION  AG083636
VERSION     AG083636.1  GI:16635438
KEYWORDS
SOURCE      Pan troglodytes male lymphoblast DNA, clone lib: PTB Chimpanzee Male
            BAC library clone: PTB-081C06.R.
ORGANISM    Pan troglodytes
REFERENCE   1  Fujiyama, A., Hattori, M., Toyoda, A., Taylor, T.D., Yada, T.,
            Tokoki, Y., Watanabe, H. and Sakaki, Y.
            BAC end sequences of library PTB
            Unpublished
            2 (bases 1 to 919)
            Direct Submission
            Submitted (02-AUG-2001) Asao Fujiyama, The Institute of Physical
            and Chemical Research (RIKEN), Genomic Sciences Center (GSC);
            1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
            (E-mail: chimpbes@sc.riken.go.jp, URL: http://hnp.gsc.riken.go.jp/,
            Tel: 81-45-503-9111, Fax: 81-45-503-9170)
            Clones are derived from the chimpanzee BAC library PTB. This BAC end
            was generated during the Rad process and may have higher chance of
            clone tracking errors.
            PRIMERS
            Sequencing: M13Rev
            LIBRARY
            Vector      : pKS145
            R.Site 1   : SacI
            R.Site 2   : SacI.
FEATURES
    source
        1..919
            location/Qualifiers
            /organism="Pan troglodytes"
            /db_xref="taxon:9598"
            /clone="PTB-081C06.R"
            /sex="male"
            /cell_type="lymphoblast"
            /clone_lib="PTB Chimpanzee Male BAC library"
BASE COUNT      202 a      205 c      193 g      292 t      27 others
ORIGIN
Alignment Scores:
Pred. No.:      501      Length:      919
Score:          48.00      Matches:      8
Percent Similarity: 81.82%      Conservative: 1
Best Local Similarity: 72.73%      Mismatches: 2
Query Match:     53.33%      Indels:      0
DB:              17      Gaps:      0

US-10-020-441-3 (1-16) x AG083636 (1-919)

```

```

Qy      2  GlyTYrGlyArgAspAsnAspArgAspPro 12
Db      729  GGGTTCGGCTGCGATGACACCGACCGACCC 761

RESULT 30
LOCUS      B0704202      980 bp      mRNA      linear      EST 16-JUL-2002
DEFINITION Bn01_04e15.A
            Bn01_AAFc ECoRC transgenic Brassica napus overexpressing BNCB17 co
            nstitutively frost_tolerant Brassica napus cDNA clone Bn01_04e15,
            mRNA sequence.
ACCESSION  B0704202
VERSION     B0704202.1  GI:21843621
KEYWORDS
SOURCE      Brassica napus
            Eukaryote; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            Rosidae; eurosids II; Brassicales; Brassicaceae; Brassica.
            1 (bases 1 to 980)
            Chagnon, J., Allard, G., Tinker, N., Robert, L., Lacroix, C., De Moors, A.,
            Singh, J., Farah, S., Courroux, P. and Hattori, J.
            Expressed Sequence Tags from constitutively frost tolerant
            transgenic Brassica napus overexpressing BNCB17
            Unpublished (2002)
            Contact: Singh, J.A.
            Eastern Cereal and Oilseed Research Centre
            Agriculture and Agri-Food Canada
            K1V 9X7 Bldg., Central Experimental Farm, Ottawa, Ontario, K1A
            0C6, Canada
            Tel: (613) 759-1662
            Fax: (613) 759-1701
            Email: singhja@em.agr.ca.
FEATURES
    source
        1..980
            /organism="Brassica napus"
            /cultivar="Westar"
            /db_xref="taxon:3708"
            /clone="Bn01_04e15"
            /clone_lib="Bn01 AAFc ECoRC transgenic Brassica napus over
            expressing BNCB17 constitutively_frost_tolerant"
            /tissue_type="fourth leaf"
            /dev_stage="3 weeks seedling grown at room temperature"
            /note="Vector: Bluescript SK+/XhoI-EcoRI; Site 1: EcoRI;
            Site 2: XhoI; Germinated in soil flats and seedlings grown
            for 3 weeks in a Conviron E-15 cabinet set at 20°C/16 hr
            light (250 Em-2sec-1) and 16 °C/ 8 hr dark. Fourth leaves
            collected at 9 am and immediately frozen."
BASE COUNT      267 a      216 c      278 g      218 t      1 others
ORIGIN
Alignment Scores:
Pred. No.:      543      Length:      980
Score:          48.00      Matches:      8
Percent Similarity: 78.57%      Conservative: 3
Best Local Similarity: 57.14%      Mismatches: 3
Query Match:     53.33%      Indels:      0
DB:              14      Gaps:      0

US-10-020-441-3 (1-16) x B0704202 (1-980)

Qy      3  TyrcGlyArgAspAsnAspArgAspProSerArgLysAsn 16
Db      584  TATGCCCGCATGATGACGATGATGATAAAGCGCAGAAC 625

RESULT 31
LOCUS      CNS06LC1      1073 bp      DNA      linear      GSS 17-JUN-2001
DEFINITION T7 end of clone ATOA012D04 of library ATOA from strain CBS 4311
            of Saccharomyces servazzii, genomic survey sequence.
ACCESSION  AL403991
VERSION     AL403991.1  GI:12164947
KEYWORDS    GSS.

```

SOURCE Saccharomyces servazzii.  
ORGANISM Saccharomyces servazzii  
REFERENCE Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
AUTHORS Saccharomycetales; Saccharomycetaceae; Saccharomyces.  
1 (bases 1 to 1073)  
Soulcet,J.L., Aigle,M., Attiguenave,F., Blandin,G.,  
Bolotin-Fukuhara,M., Bon,E., Broctier,P., Casaregola,S.,  
deMontigny,J., Dujon,B., Durenne,P., Lepingle,A., Llorente,B.,  
Malpertuy,A., Neuvéglise,C., Ozier-Kalogeropoulos,O., Potier,S.,  
Saurin,W., Tekala,F., Toffano-Nioche,C., Wesolowski-Louvel,M.,  
Wincker,P. and Weissenbach,J.  
TITLE Genomic exploration of the hemiascomycetous yeasts: 1. A set of  
JOURNAL yeast species for molecular evolution studies  
MEDLINE FEMS Lett. 487 (1), 3-12 (2000)  
PUBMED 20584711  
1152876  
2 (bases 1 to 1073)  
Casaregola,S., Lepingle,A., Bon,E., Neuvéglise,C., Nguyen,H.,  
Attiguenave,F., Wincker,P. and Galliardin,C.  
REFERENCE Genomic exploration of the hemiascomycetous yeasts: 7.  
AUTHORS Saccharomyces servazzii  
TITLE FEMS Lett. 487 (1), 47-51 (2000)  
JOURNAL 20584717  
MEDLINE 1152882  
PUBMED 3 (bases 1 to 1073)  
REFERENCE Genoscope.  
AUTHORS Direct Submission  
TITLE Submitted (07-SEP-2000) Genoscope - Centre National de Séquençage,  
JOURNAL 2 rue Gaston Crémieux, CP 5706, 91057 EVRY cedex, FRANCE. (E-mail :  
seqref@genoscope.cns.fr - Web : www.genoscope.cns.fr)  
COMMENT This GSS is part of a random genomic sequencing program of thirteen  
yeast species: Saccharomyces bayanus var. uvarum, Saccharomyces  
exiguus, Saccharomyces servazzii, Zygosaccharomyces rouxii,  
Saccharomyces kluyveri, Kluyveromyces thermotolerans, Kluyveromyces  
lactis var. lactis, Kluyveromyces marxianus var. marxianus, Pichia  
agustae, Debaryomyces hansenii var. hansenii, Pichia sorbitophila,  
Candida tropicalis and Yarrowia lipolytica. Genomic inserts of 3 to  
5 kb were prepared and both extremities were sequenced. See  
keywords for description of this sequence and for the sequence of  
the other extremity of this insert.  
FEATURES  
source Location/Qualifiers  
1..1073  
/organism="Saccharomyces servazzii"  
/strain="CBS 4311"  
/db\_xref="taxon:27293"  
/clone="AT0AA012D04"  
/clone\_1lb="AT0AA"  
/note="end : T7"  
2..>262  
/note="similar to Saccharomyces cerevisiae ORF YER052c [  
HOM3 : L-aspartate 4-P-transferase ]"  
/evidence="not experimental"  
misc\_feature  
BASE COUNT 380 a 176 c 162 g 349 t 6 others  
ORIGIN

Alignment Scores:  
Pred. No.: 607 Length: 1073  
Score: 48.00 Matches: 9  
Percent Similarity: 68.75% Conservative: 2  
Best Local Similarity: 56.25% Mismatches: 5  
Query Match: 53.33% Indels: 0  
DB: 17 Gaps: 0  
US-10-020-441-3 (1-16) x CNS061C1 (1-1073)  
Qy 1 ValGlyTyrGlyArgAspAsnAspArgAspProSerArgLysAsn 16  
Db 336 GTTGCGCATCAAGAGATGACCAAAATGTATCATCCGTAACAAAAAAC 289  
RESULT 32  
LOCUS AU184369  
DEFINITION AU184369 Rice root Oryza sativa (japonica cultivar-group) cDNA

ACCESSION clone R2083, mRNA sequence.  
VERSION AU184369  
KEYWORDS AU184369.1 GI:14192158  
SOURCE EST.  
ORGANISM Oryza sativa (japonica cultivar-group).  
REFERENCE Oryza sativa (japonica cultivar-group).  
AUTHORS Eukaryota; Viridiplantae; Embryophyta; Tracheophyta;  
TITLE Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;  
JOURNAL Ehrhartoideae; Oryzaceae; Oryza.  
1 (bases 1 to 446)  
Sasaki,T. and Yamamoto,K.  
REFERENCE Rice cDNA from root (2001)  
AUTHORS Unpublished (2001)  
TITLE Contact: Takuji Sasaki  
JOURNAL National Institute of Agrobiological Resources  
COMMENT Rice Genome Research Program, Kamondai 2-1-2, Tsukuba, Ibaraki  
305-8602, Japan  
Tel: 81-298-38-7441  
Fax: 81-298-38-7468  
Email: tsasaki@abr.affrc.go.jp, URL: http://ryg.dna.affrc.go.jp/  
PROJECT = "RGP".  
FEATURES  
source Location/Qualifiers  
1..446  
/organism="Oryza sativa (japonica cultivar-group)"  
/cultivar="Nipponbare"  
/db\_xref="taxon:39947"  
/clone="R2083"  
/clone\_1lb="Rice root"  
/note="Prepared from seedling root. "  
BASE COUNT 125 a 92 c 105 g 123 t 1 others  
ORIGIN

Alignment Scores:  
Pred. No.: 249 Length: 446  
Score: 47.50 Matches: 10  
Percent Similarity: 64.71% Conservative: 1  
Best Local Similarity: 58.82% Mismatches: 3  
Query Match: 52.78% Indels: 3  
DB: 9 Gaps: 1  
US-10-020-441-3 (1-16) x AU184369 (1-446)  
Qy 2 GlyTyrGlyArg-----AspAsnAspArgAspProSerArgLys 15  
Db 63 GGCTACGGCCGCAAGAAACATGACATGACTCTATGATGAAAGAG 113  
RESULT 33  
LOCUS AU029604  
DEFINITION 447 bp mRNA linear EST 01-APR-2002  
ACCESSION AU029604 Rice panicle shorter than 3cm Oryza sativa (japonica  
KEYWORDS cultivar-group) cDNA clone E31136\_6Z, mRNA sequence.  
SOURCE AU029604.1 GI:3762852  
ORGANISM EST.  
REFERENCE Oryza sativa (japonica cultivar-group).  
AUTHORS Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
TITLE Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;  
JOURNAL Ehrhartoideae; Oryzaceae; Oryza.  
1 (bases 1 to 447)  
Sasaki,T. and Yamamoto,K.  
REFERENCE Rice cDNA from panicle  
AUTHORS Unpublished (1997)  
TITLE Contact: Takuji Sasaki  
JOURNAL National Institute of Agrobiological Resources  
COMMENT Rice Genome Research Program, Kamondai 2-1-2, Tsukuba, Ibaraki  
305-8602, Japan  
Tel: 81-298-38-7441  
Fax: 81-298-38-7468  
Email: tsasaki@abr.affrc.go.jp, URL: http://ryg.dna.affrc.go.jp/  
PROJECT = "RGP".  
FEATURES  
source Location/Qualifiers  
POLYA=No.



COMMENT	Contact: Dong HT Laboratory of Functional Genetics Bio-technology Institute of Zhejiang University Kaixuan Road 268#, Hangzhou, Zhejiang, P.R.China Tel: 0086-571-86892051 Fax: 0086-571-86961525 Email: htdong@zjuem.zju.edu.cn Seq primer: M13 forward primer.				
FEATURES	Location/Qualifiers				
SOURCE	1..480 /organism="Oryza sativa" /db_xref="taxon:4530" /clone="F008C10" /clone_lib="Oryza sativa mature leaf library induced by M.grisea" /tissue_type="leaf" /dev_stage="Mature stage" /note="Vector: pSPori2"				
BASE COUNT	134 a	99 c	114 g	133 t	
ORIGIN					
Alignment Scores:					
Pred. No.:	273	Length:	480		
Score:	47.50	Matches:	10		
Percent Similarity:	64.71%	Conservative:	1		
Best Local Similarity:	58.82%	Mismatches:	3		
Query Match:	52.78%	Indels:	3		
DB:	13	Gaps:	1		
US-10-020-441-3 (1-16) X BI809420 (1-480)					
OY 2 G1YTYGYG1YAG-----AphAphAAnaPaPaGApProSeRaRg1Ys 15					
Db 96 GGCTACGGCCGACAGAAACATGATGACATGACTCGATGATGAGAAAGAAG 146					
RESULT 37					
BI306405/c					
LOCUS	BI306405	497 bp	mRNA	linear	
DEFINITION	NL_4_114 Drought stress (leaf) Oryza sativa cDNA clone NL_4_114 3',				
ACCESSION	BI306405	mRNA sequence.			
VERSION	BI306405.1	GI:14981727			
KEYWORDS	EST.				
SOURCE	Oryza sativa.				
ORGANISM	Oryza sativa				
REFERENCE	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;				
AUTHORS	Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;				
TITLE	Ehhartriidae; Oryzae; Oryza.				
JOURNAL	1 (bases 1 to 497)				
COMMENT	Reddy,A.R., Ramakrishna,W., Chandrasekhar,A., Nagabhushan,I., Rayindrababu,P. and Bennetzen,J.L. Novel EST enrichment with normalized cDNA libraries from drought stressed rice (Oryza sativa L.cv Nagina 22) Unpublished (2001)				
CONTACT	Contact: Reddy AR				
DEPARTMENT	Department of Plant Sciences, School of Life Sciences				
UNIVERSITY	University of Hyderabad				
P.O.	Central University, Hyderabad-500 046, A.P, India				
TEL:	0091-40-3010265				
FAX:	0091-40-3010145				
EMAIL:	arjule@uohyd.ernet.in				
INSERT LENGTH:	497	STD ERROR:	0.00		
PLATE:	4	ROW:	1	COLUMN:	14
SEQ PRIMER:	GTAAACGACGGCCGATG.				
FEATURES	Location/Qualifiers				
SOURCE	1..497 /organism="Oryza sativa" /cultivar="Nagina 22 (indica sub sp)" /db_xref="taxon:4530" /clone="NL_4_114" /clone_lib="Drought stress (leaf)" /tissue_type="mature leaf tissue" /dev_stage="35 day-old seedlings"				

```

/Note=Organ: leaf; Vector: TT73Pac; ESTs from normalized
leaf cDNA library from drought stressed seedlings "
BASE COUNT      134 a      134 c      106 g      123 t
ORIGIN

Alignment Scores:
Pred. No.:      285                      Length:      497
Score:          47.50                   Matches:      10
Percent Similarity: 64.71%              Conservative: 1
Best Local Similarity: 58.82%            Mismatches:   3
Query Match:    52.78%                   Indels:       3
DB:             13                       Gaps:         1

US-10-020-441-3 (1-16) x B1306405 (1-497)

Oy      2 G1YTYrG1YArG-----AspAspAsnAspArGaSpProSeArGlyls 15
Db      366 GGCCTACGGCCGCAAGAAACATGATGACATGACTCTGTATGTGAAAAGAG 316

RESULT 38
LOCUS      BM419755                      531 bp      mRNA      linear      EST 28-JAN-2002
DEFINITION R018C09 Oryza sativa mature leaf library induced by M.grisea Oryza
sativa cDNA clone R018C09, mRNA sequence.
ACCESSION  BM419755
VERSION     BM419755.1  GI:18386557
KEYWORDS   EST.
SOURCE     Oryza sativa.
ORGANISM   Oryza sativa.
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
            Ehrhartoideae; Oryzaceae; Oryza.
REFERENCE   1 (bases 1 to 531)
AUTHORS     Dong,H.T., Li,D.B., Zhuang,X.F., Dai,C.G., Sun,L.X., Pei,Y.X., Wu
            ,H.F., Jiang,Y.X., Yu,F.C., Gao,Q.K. and Lou,Y.C.
            A Gene Expression screen in Oryza sativa
            Unpublished (2001)
TITLE       Contact: Dong HT
            Laboratory of Functional Genetics
            Bio-technology Institute of Zhejiang University
            Kaixuan Road 268#, Hangzhou, Zhejiang, P.R.China
            Tel: 0086-571-86892051
            Fax: 0086-571-86961525
            Email: htdong@zjuem.zju.edu.cn
            Seg primer: M13 forward primer.
FEATURES
            Location/Qualifiers
                source          1..531
                                /organism="Oryza sativa"
                                /db_xref="taxon:4530"
                                /clone="R018C09"
                                /clone_id="Oryza sativa mature leaf library induced by
                                M.grisea"
                                /tissue_type="leaf"
                                /dev_stage="Mature stage"
                                /note="Vector: pSport2"
BASE COUNT      144 a      111 c      133 g      143 t
ORIGIN

Alignment Scores:
Pred. No.:      309                      Length:      531
Score:          47.50                   Matches:      10
Percent Similarity: 64.71%              Conservative: 1
Best Local Similarity: 58.82%            Mismatches:   3
Query Match:    52.78%                   Indels:       3
DB:             13                       Gaps:         1

US-10-020-441-3 (1-16) x BM419755 (1-531)

Oy      2 G1YTYrG1YArG-----AspAspAsnAspArGaSpProSeArGlyls 15
Db      146 GGCCTACGGCCGCAAGAAACATGATGACATGACTCTGTATGTGAAAAGAG 196

RESULT 39

```

AU082192 AU082192 713 bp mRNA linear EST 02-APR-2002  
 LOCUS AU082192 Rice callus Oryza sativa (japonica cultivar-group) cDNA  
 DEFINITION clone C10261, mRNA sequence.  
 ACCESSION AU082192 GI:6799241  
 VERSION AU082192.1 GI:6799241  
 SOURCE EST.  
 ORGANISM Oryza sativa (japonica cultivar-group).  
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 Spermatophyta; Magnoliophyta, Liliopsida; Poales; Poaceae;  
 Eriactoidae; Oryzaceae; Oryza.  
 1 (bases 1 to 713)  
 Sasaki, T. and Yamamoto, K.  
 Rice cDNA from callus (2000)  
 Unpublished (2000)  
 CONTACT Takuji Sasaki  
 National Institute of Agrobiological Resources  
 Rice Genome Research Program, Kannondai 2-1-2, Tsukuba, Ibaraki  
 305-8602, Japan  
 Tel: 81-298-38-7441  
 Fax: 81-298-38-7468  
 Email: tsasaki@abrr.affrc.go.jp, URL: http://rgp.dna.affrc.go.jp/  
 PROJECT = 'RGP'  
 Location/Qualifiers  
 1..713  
 /organism="Oryza sativa (japonica cultivar-group)"  
 /cultivar="Nipponbare"  
 /db\_xref="taxon:39947"  
 /clone="C10261"  
 /clone\_1db="Rice callus"  
 /note="Vector: pBluescript II SK+; Site 1: SalI; Site 2:  
 NotI; cDNA prepared from rice callus mRNAs by using  
 oligo(dT) as a primer and ligating to the SalI-NotI site  
 of pBluescript II SK+ phagemid." 4 others

BASE COUNT 190 a 136 c 210 g 173 t  
 ORIGIN

Alignment Scores:  
 Pred. No.: 445 Length: 713  
 Score: 47.50 Matches: 10  
 Percent Similarity: 64.71% Conservative: 1  
 Best Local Similarity: 58.82% Mismatches: 3  
 Query Match: 52.78% Indels: 3  
 DB: 9 Gaps: 1

US-10-020-441-3 (1-16) x AU082192 (1-713)

Cy 2 G|YTYRG|YARG-----AspAspAsnAspArgAspProSerArgLys 15  
 |||||  
 Db 330 GGCTACGGCCGCGAGAAACATGATGACATGACTGATGATGATGAAGAAG 380  
 |||||

RESULT 40  
 A1009417 275 bp mRNA linear EST 08-JAN-1999  
 LOCUS EST203868 Normalized rat heart, Bento Soares Rattus sp. cDNA clone  
 DEFINITION RHBJ70 3' end, mRNA sequence.  
 ACCESSION A1009417  
 VERSION A1009417.1 GI:4133037  
 KEYWORDS EST.  
 SOURCE Rattus sp.  
 ORGANISM Rattus sp.  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;  
 Rattus.  
 1 (bases 1 to 275)  
 Lee, N.H., Glodek, A., Chandra, I., Mason, T.M., Quackenbush, J.,  
 Kerlavage, A.R., and Adams, M.D.  
 Rat Genome Project: Generation of a Rat EST (RST) Catalog & Rat  
 Gene Index  
 Unpublished (1998)  
 On Jun 15, 1998 this sequence version replaced gi:3223249.  
 CONTACT: Lee, NH

The Institute for Genomic Research  
 9712, Medical Center Drive, Rockville, MD 20850, USA  
 Tel: (301)-838-3529  
 Fax: (301)-838-0208  
 Email: nhlee@tigr.org  
 Seq primer: M13-21.  
 Location/Qualifiers  
 1..275  
 /organism="Rattus sp."  
 /db\_xref="taxon:10118"  
 /clone="RHBJ70"  
 /clone\_1db="Normalized rat heart, Bento Soares"  
 /note="Organ: heart; Vector: pT73pac; Site\_1: EcoRI;  
 Site\_2: NotI"

BASE COUNT 75 a 55 c 75 g 70 t  
 ORIGIN

Alignment Scores:  
 Pred. No.: 166 Length: 275  
 Score: 47.00 Matches: 8  
 Percent Similarity: 66.67% Conservative: 2  
 Best Local Similarity: 53.33% Mismatches: 5  
 Query Match: 52.22% Indels: 0  
 DB: 9 Gaps: 0

US-10-020-441-3 (1-16) x A1009417 (1-275)

Cy 1 ValG|TYRG|YARGAspAspAsnAspArgAspProSerArgLys 15  
 |||||  
 Db 20 ATTGGATTGGAGGCAAGAGTGGAGGAGGAGACCCGACAGGAAA 64  
 |||||

Search completed: April 8, 2003, 19:52:14  
 Job time : 229.557 secs

GenCore version 5.1.4 p5 4578  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: April 8, 2003, 18:35:44 / Search time 8.13115 Seconds  
(without alignments)  
603.461 Million cell updates/sec

Title: US-10-020-441-3  
Perfect score: 90  
Sequence: 1 VGYGRDNDNDPSRKN 16

Scoring table:  
BLOSUM62  
Xgapop 10.0, Xgapext 0.5  
Ygapop 10.0, Ygapext 0.5  
Fgapop 6.0, Fgapext 7.0  
Delop 6.0, Delext 7.0

Searched: 441363 seqs, 153338381 residues  
Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Command line parameters:  
-MODEL=frame+ p2n.model -DEV=xlh  
-Q=/cgn2\_1/USPTO.epool/US10020441/runat\_02042003\_093013\_21004/app\_query.fasta-1.462  
-DB=Issued\_Patents\_NA -QEMT=faastap -SUFFIX=rni -MINMATCH=0.1 -LOOPCL=0  
-LOOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=bl0sum62 -TRANS=human40.cdi  
-LIST=45 -DOCALIGN=200 -THR SCORE=DCT -THR MAX=100 -THR MIN=0 -ALIGN=40  
-MODE=LOCAL -OUTFMT=ptco -NOM=ext -HEAPSIZE=500 -MITLEN=0 -MAXLEN=200000000  
-USRR=US10020441 @CGN 1.1 47 @runat\_02042003\_093013\_21004 -NCPU=6 -ICPU=3  
-NO\_XLPRY -NO\_MMAT -LARGEQUERY -NEG\_SCORES=0 -WAIT -LONGLOG -DEV\_TIMEOUT=120  
-WARN\_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -Fgapop=6 -Fgapext=7  
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Issued Patents NA:\*

- 1: /cgn2\_6/ptodata/1/ina/5A.COMB.seq:\*
- 2: /cgn2\_6/ptodata/1/ina/5B.COMB.seq:\*
- 3: /cgn2\_6/ptodata/1/ina/6A.COMB.seq:\*
- 4: /cgn2\_6/ptodata/1/ina/6B.COMB.seq:\*
- 5: /cgn2\_6/ptodata/1/ina/PTUS.COMB.seq:\*
- 6: /cgn2\_6/ptodata/1/ina/backfile1.seq:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
C 1	52	57.8	2833	4 US-09-276-531-23	Sequence 23, Appli
C 2	52	57.8	2880	2 US-08-987-289-1	Sequence 1, Appli
3	44	48.9	1842	4 US-09-832-498-1	Sequence 1, Appli
4	44	48.9	1842	4 US-09-832-498-1	Sequence 1, Appli
5	43	47.8	1777	4 US-09-134-001C-1927	Sequence 1927, Ap
C 6	43	47.8	1149	3 US-08-660-645A-8	Sequence 8, Appli
C 7	43	47.8	1149	3 US-09-298-718-8	Sequence 8, Appli
C 8	43	47.8	1149	4 US-09-546-969-8	Sequence 8, Appli
9	43	47.8	8625	4 US-08-980-832-1	Sequence 1, Appli
10	43	47.8	11233	4 US-08-980-832-27	Sequence 27, Appli
C 11	42.5	47.2	8520	4 US-09-272-032-7	Sequence 7, Appli
12	42	46.7	2372	1 US-07-903-103-1	Sequence 1, Appli

13	42	46.7	2372	1 US-08-044-619A-1	Sequence 1, Appli
14	42	46.7	2372	1 US-08-283-911-1	Sequence 2, Appli
15	42	46.7	2372	1 US-08-245-500A-2	Sequence 2, Appli
16	42	46.7	2372	1 US-08-390-546-2	Sequence 2, Appli
17	42	46.7	2372	1 US-08-390-479A-2	Sequence 2, Appli
18	42	46.7	2372	1 US-08-557-393-2	Sequence 2, Appli
19	42	46.7	2372	1 US-08-390-516C-2	Sequence 2, Appli
20	42	46.7	2372	1 US-08-390-517A-2	Sequence 2, Appli
21	42	46.7	2372	1 US-08-390-515A-2	Sequence 2, Appli
22	42	46.7	2372	1 US-08-801-718-2	Sequence 2, Appli
23	42	46.7	2372	3 US-09-073-567-1	Sequence 1, Appli
24	42	46.7	2372	4 US-09-280-805-1	Sequence 1, Appli
25	42	46.7	2372	4 US-09-048-810-1	Sequence 1, Appli
26	42	46.7	2372	4 US-09-170-159A-2	Sequence 2, Appli
27	42	46.7	2372	4 US-09-480-718-43	Sequence 43, Appli
C 28	41	45.6	312	3 US-09-106-486-3	Sequence 3, Appli
C 29	41	45.6	312	4 US-09-377-446-3	Sequence 3, Appli
C 30	41	45.6	504	4 US-09-512-342-19	Sequence 19, Appli
C 31	41	45.6	670	4 US-09-221-017B-63	Sequence 63, Appli
C 32	41	45.6	1496	3 US-09-048-889-9	Sequence 9, Appli
C 33	41	45.6	1546	4 US-08-796-899-22	Sequence 22, Appli
C 34	41	45.6	1881	4 US-08-434-288-5	Sequence 5, Appli
C 35	41	45.6	7266	3 US-08-336-308A-9	Sequence 9, Appli
C 36	41	45.6	7266	3 US-08-822-324-5	Sequence 5, Appli
C 37	41	45.6	7266	3 US-09-490-931-9	Sequence 9, Appli
C 38	41	45.6	8439	4 US-09-221-017B-473	Sequence 473, App
C 39	41	45.6	8640	1 US-08-570-311-28	Sequence 28, Appli
40	41	45.6	38564	3 US-09-734-673-3	Sequence 3, Appli
41	41	45.6	72928	3 US-09-009-913-1	Sequence 1, Appli
C 42	41	45.6	4411529	4 US-09-103-840A-1	Sequence 1, Appli
C 43	40.5	45.0	5057	3 US-08-651-999A-6	Sequence 6, Appli
C 44	40.5	45.0	5057	4 US-09-385-752-6	Sequence 6, Appli
45	40	44.4	153	2 US-08-679-865-41	Sequence 41, Appli

ALIGNMENTS

RESULT 1  
US-09-276-531-23/C  
Sequence 23, Application US/09276531  
Patent No. 6183968  
GENERAL INFORMATION:  
APPLICANT: Bandman, Olga  
APPLICANT: Lal, Preeti  
APPLICANT: Hillman, Jennifer L.  
APPLICANT: Yue, Henry  
APPLICANT: Reddy, Koopa  
APPLICANT: Guegler, Karl J.  
APPLICANT: Baughn, Mariah R.  
TITLE OF INVENTION: COMPOSITION FOR THE DETECTION OF GENES ENCODING  
TITLE OF INVENTION: RECEPTORS AND PROTEINS ASSOCIATED WITH CELL PROLIFERATION  
NUMBER OF SEQUENCES: 134  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: INCYTE PHARMACEUTICALS, INC.  
STREET: 3174 PORTER DRIVE  
CITY: PALO ALTO  
STATE: CALIFORNIA  
COUNTRY: USA  
ZIP: 94304  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Word Perfect 6.1 for Windows/MS-DOS 6.2  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/276,531  
FILING DATE: Herewith  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 60/079,677  
FILING DATE: March 27, 1998  
CLASSIFICATION:  
ATTORNEY/AGENT INFORMATION:

```

; NAME: Lynn E. Murry, Ph.D.
; REGISTRATION NUMBER: 42,918
; REFERENCE/DOCKET NUMBER: PA-0008 US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (650) 855-0555
; TELEFAX: (650) 845-4166
; INFORMATION FOR SEQ ID NO: 23:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2833 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; LIBRARY: FIBRANT01
; CLONE: 150629
; US-09-276-531-23

Alignment Scores:
Pred. No.: 7.45 Length: 2833
Score: 52.00 Matches: 11
Percent Similarity: 85.71% Conservative: 1
Best Local Similarity: 78.57% Mismatches: 1
Query Match: 57.78% Indels: 1
DB: 4 Gaps: 0

US-10-020-441-3 (1-16) x US-09-276-531-23 (1-2833)
Qy 1 ValGIYTYrGIYAArgAspAspAsnAsp-ArgAspProser 13
Db 1552 ATGGGATATGTGCTGACGACATGATCAGAGACCCAACT 1513

RESULT 2
US-08-987-289-1/c
; Sequence 1, Application US/08987289
; Patent No. 5994098
; GENERAL INFORMATION:
; APPLICANT: HU, ERDING
; APPLICANT: ZHU, YUAN
; TITLE OF INVENTION: A Human 7-TM Receptor Similar
; TITLE OF INVENTION: to Murine Frizzled-6 gene
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: RATNER & PRESTIA
; STREET: P.O. BOX 980
; CITY: VALLEY FORGE
; STATE: PA
; COUNTRY: USA
; ZIP: 19482
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FASTSEQ for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/987,289
; FILING DATE: 09-DEC-1997
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 60/048,725
; FILING DATE: 02-JUN-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: PRESTIA, PAUL F
; REGISTRATION NUMBER: 23,031
; REFERENCE/DOCKET NUMBER: GH-70060
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 610-407-0700
; TELEFAX: 610-407-0701
; TELEX: 846169
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2880 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
;
```

```

; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; US-08-987-289-1

Alignment Scores:
Pred. No.: 7.59 Length: 2880
Score: 52.00 Matches: 11
Percent Similarity: 85.71% Conservative: 1
Best Local Similarity: 78.57% Mismatches: 1
Query Match: 57.78% Indels: 1
DB: 2 Gaps: 0

US-10-020-441-3 (1-16) x US-08-987-289-1 (1-2880)
Qy 1 ValGIYTYrGIYAArgAspAspAsnAsp-ArgAspProser 13
Db 1614 ATGGGATATGTGCTGACGACATGATCAGAGACCCAACT 1575

RESULT 3
US-09-832-498-1
; Sequence 1, Application US/09832498
; Patent No. 6410291
; GENERAL INFORMATION:
; APPLICANT: Danielsen, Steffen
; APPLICANT: Schneider, Palte
; TITLE OF INVENTION: Polypeptides having haloperoxidase activity
; FILE REFERENCE: 10040.200-US
; CURRENT APPLICATION NUMBER: US/09/832,498
; CURRENT FILING DATE: 2001-04-11
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 1
; LENGTH: 1842
; TYPE: DNA
; ORGANISM: Geniculisporium sp.
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(1842)
; US-09-832-498-1

Alignment Scores:
Pred. No.: 11.1 Length: 1842
Score: 44.00 Matches: 8
Percent Similarity: 90.00% Conservative: 1
Best Local Similarity: 80.00% Mismatches: 1
Query Match: 48.89% Indels: 0
DB: 4 Gaps: 0

US-10-020-441-3 (1-16) x US-09-832-498-1 (1-1842)
Qy 5 ArgAspAspAsnAspArgAspProserArg 14
Db 1096 CGTGACGACACTTCGCGACCCCAATCGC 1125

RESULT 4
US-09-832-614A-1
; Sequence 1, Application US/09832614A
; Patent No. 6410292
; GENERAL INFORMATION:
; APPLICANT: No. 6410292zymes A/S
; APPLICANT: Danielsen, Steffen
; APPLICANT: Schneider, Palte
; TITLE OF INVENTION: Nucleic acids encoding polypeptides having haloperoxidase activity
; FILE REFERENCE: 10173.200-US
; CURRENT APPLICATION NUMBER: US/09/832,614A
; CURRENT FILING DATE: 2001-04-11
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 1
; LENGTH: 1842
; TYPE: DNA
; ORGANISM: Geniculisporium sp.
; FEATURE:
;
```



```

; NAME/KEY: CDS
; LOCATION: (1)..(1842)
; OTHER INFORMATION:
US-09-832-614A-1

Alignment Scores:
Pred. No.: 111      Length: 1842
Score: 44.00      Matches: 8
Percent Similarity: 90.00%      Conservative: 1
Best Local Similarity: 80.00%      Mismatches: 1
Query Match: 48.89%      Indels: 0
DB: 4      Gaps: 0

US-10-020-441-3 (1-16) x US-09-832-614A-1 (1-1842)

Qy 5 ArgAaPaPaPaPaPaPaPaPaPaPaPaPaPaPaPa 14
Db 1096 CGTACGACGACCTTCGCGACCCCAATCGC 1125

RESULT 5
US-09-134-001C-1927
; Sequence 1927, Application US/09134001C
; Patent No. 6380370
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
; FILE REFERENCE: GTC-007
; CURRENT APPLICATION NUMBER: US/09/134,001C
; PRIOR FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/064,964
; PRIOR FILING DATE: 1997-11-08
; PRIOR APPLICATION NUMBER: US 60/055,779
; PRIOR FILING DATE: 1997-08-14
; NUMBER OF SEQ ID NOS: 5674
; SEQ ID NO 1927
; LENGTH: 777
; TYPE: DNA
; ORGANISM: Staphylococcus epidermidis
US-09-134-001C-1927

Alignment Scores:
Pred. No.: 62.3      Length: 777
Score: 43.00      Matches: 7
Percent Similarity: 88.89%      Conservative: 1
Best Local Similarity: 77.78%      Mismatches: 1
Query Match: 47.78%      Indels: 0
DB: 4      Gaps: 0

US-10-020-441-3 (1-16) x US-09-134-001C-1927 (1-777)

Qy 2 GlyTyTgLyArgaPaPaPaPaPaPaPaPaPa 10
Db 547 GGTATGTGTATATGATGATGATGATGATG 573

RESULT 6
US-08-660-645A-8/C
; Sequence 8, Application US/08660645A
; Patent No. 6087152
; GENERAL INFORMATION:
; APPLICANT: Hohmann, Hans-Peter
; APPLICANT: Pasamontes, Luis
; APPLICANT: Tessier, Michel
; APPLICANT: van Loon, Adolphus
; TITLE OF INVENTION: FERMENTATIVE CAROTENOID PRODUCTION
; NUMBER OF SEQUENCES: 47
; CORRESPONDENCE ADDRESSES:
; ADDRESSEE: Hoffmann-La Roche Inc.
; STREET: 340 Kingsland Street
; CITY: Nutley
; STATE: NJ
; COUNTRY: USA
; ZIP: 07110
```

```

; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/660,645A
; FILING DATE: 07-JUN-1996
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: EP 95108888.9
; FILING DATE: 09-JUN-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Pokras, Bruce A.
; REGISTRATION NUMBER: 32,748
; REFERENCE/DOCKET NUMBER: RAN 6002/170
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (201) 235-5801
; TELEFAX: (201) 235-2363
; INFORMATION FOR SEQ ID NO: 8:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1149 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
US-08-660-645A-8

Alignment Scores:
Pred. No.: 96.9      Length: 1149
Score: 43.00      Matches: 9
Percent Similarity: 69.23%      Conservative: 0
Best Local Similarity: 69.23%      Mismatches: 4
Query Match: 47.78%      Indels: 0
DB: 3      Gaps: 0

US-10-020-441-3 (1-16) x US-08-660-645A-8 (1-1149)

Qy 2 GlyTyTgLyArgaPaPaPaPaPaPaPaPaPa 14
Db 925 GGAAGCGGTGCGATCGCGCGATGATGATGATGATG 887

RESULT 7
US-09-298-718-8/C
; Sequence 8, Application US/09298718
; Patent No. 6124113
; GENERAL INFORMATION:
; APPLICANT: Hohmann, Hans-Peter
; APPLICANT: Pasamontes, Luis
; APPLICANT: Tessier, Michel
; APPLICANT: van Loon, Adolphus
; TITLE OF INVENTION: FERMENTATIVE CAROTENOID PRODUCTION
; NUMBER OF SEQUENCES: 47
; CORRESPONDENCE ADDRESSES:
; ADDRESSEE: Hoffmann-La Roche Inc.
; STREET: 340 Kingsland Street
; CITY: Nutley
; STATE: NJ
; COUNTRY: USA
; ZIP: 07110
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/298,718
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/660,645
; ATTORNEY/AGENT INFORMATION:
```

NAME: Pokras, Bruce A.  
REGISTRATION NUMBER: 32,748  
REFERENCE/DOCKET NUMBER: RAN 6002/170  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (201) 235-5801  
TELEFAX: (201) 235-2363  
INFORMATION FOR SEQ ID NO: 8:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1149 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
US-09-298-718-8

Alignment Scores:  
Pred. No.: 96.9 Length: 1149  
Score: 43.00 Matches: 9  
Percent Similarity: 69.23% Conservative: 0  
Best Local Similarity: 69.23% Mismatches: 4  
Query Match: 47.78% Indels: 0  
DB: 3 Gaps: 0

US-10-020-441-3 (1-16) x US-09-298-718-8 (1-1149)  
Qy 2 G1YTYG1YARGAspAspAsnAspArgAspProSerArg 14  
Db 925 GGAAGCGGTCCGATCCGCGCATCGATGAGCCCAAGCCGC 887

RESULT 8  
US-09-546-969-8/c  
Sequence 8, Application US/09546969  
Patent No. 6207409  
GENERAL INFORMATION:  
APPLICANT: Hohmann, Hans-Peter  
APPLICANT: Pasamontes, Luis  
APPLICANT: Tessier, Michel  
APPLICANT: van Loon, Adolphus  
TITLE OF INVENTION: FERMENTATIVE CAROTENOID PRODUCTION  
NUMBER OF SEQUENCES: 47  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Hoffmann-La Roche Inc.  
STREET: 340 Kingsland Street  
CITY: Nutley  
STATE: NJ  
COUNTRY: USA  
ZIP: 07110  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/546,969  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/660,645  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: Pokras, Bruce A.  
REGISTRATION NUMBER: 32,748  
REFERENCE/DOCKET NUMBER: RAN 6002/170  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (201) 235-5801  
TELEFAX: (201) 235-2363  
INFORMATION FOR SEQ ID NO: 8:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1149 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)

US-09-546-969-8

Alignment Scores:  
Pred. No.: 96.9 Length: 1149  
Score: 43.00 Matches: 9  
Percent Similarity: 69.23% Conservative: 0  
Best Local Similarity: 69.23% Mismatches: 4  
Query Match: 47.78% Indels: 0  
DB: 4 Gaps: 0

US-10-020-441-3 (1-16) x US-09-546-969-8 (1-1149)  
Qy 2 G1YTYG1YARGAspAspAsnAspArgAspProSerArg 14  
Db 925 GGAAGCGGTCCGATCCGCGCATCGATGAGCCCAAGCCGC 887

RESULT 9  
US-08-980-832-1  
Sequence 1, Application US/08980832B  
Patent No. 6291204  
GENERAL INFORMATION:  
APPLICANT: Pasamontes, Luis  
APPLICANT: Tsyganov, Yuri  
TITLE OF INVENTION: Improved Fermentative Carotenoid Production  
FILE REFERENCE: Improved Fermentative Carotenoid  
CURRENT APPLICATION NUMBER: US/08/980,832B  
CURRENT FILING DATE: 1997-12-01  
NUMBER OF SEQ ID NOS: 66  
SOFTWARE: Patentin Ver. 2.1  
SEQ ID NO 1  
LENGTH: 8625  
TYPE: DNA  
ORGANISM: Flavobacterium sp. R1534  
FEATURE:  
NAME/KEY: unsure  
LOCATION: (8348)..(8349)  
FEATURE:  
NAME/KEY: unsure  
LOCATION: (8539)..(8540)  
FEATURE:  
NAME/KEY: unsure  
LOCATION: (8581)  
FEATURE:  
NAME/KEY: unsure  
LOCATION: (8590)  
FEATURE:  
NAME/KEY: unsure  
LOCATION: (8592)  
FEATURE:  
NAME/KEY: unsure  
LOCATION: (8602)..(8604)  
US-08-980-832-1

Alignment Scores:  
Pred. No.: 940 Length: 8625  
Score: 43.00 Matches: 9  
Percent Similarity: 69.23% Conservative: 0  
Best Local Similarity: 69.23% Mismatches: 4  
Query Match: 47.78% Indels: 0  
DB: 4 Gaps: 0

US-10-020-441-3 (1-16) x US-08-980-832-1 (1-8625)  
Qy 2 G1YTYG1YARGAspAspAsnAspArgAspProSerArg 14  
Db 6018 GGAAGCGGTCCGATCCGCGCATCGATGAGCCCAAGCCGC 6056

RESULT 10  
US-08-980-832-27  
Sequence 27, Application US/08980832B  
Patent No. 6291204  
GENERAL INFORMATION:  
APPLICANT: Pasamontes, Luis

```

      APPLICANT: Tsuygan'kov, Yuri
      TITLE OF INVENTION: Improved Fermentative Carotenoid Production
      FILE REFERENCE: Improved Fermentative Carotenoid
      CURRENT APPLICATION NUMBER: US/08/980,832B
      NUMBER OF SEQ ID NOS: 66
      SOFTWARE: PatentIn Ver. 2.1
      SEQ ID NO 27
      LENGTH: 11233
      TYPE: DNA
      ORGANISM: Unknown
      FEATURE:
      OTHER INFORMATION: Description of Unknown Organism: Plasmid pzea4
US-08-980-832-27

Alignment Scores:
Pred. No.: 1.27e+03      Length: 11233
Score: 43.00      Matches: 9
Percent Similarity: 69.23%      Conservative: 0
Best Local Similarity: 69.23%      Mismatches: 4
Query Match: 47.78%      Indels: 0
DB: 4      Gaps: 0

US-10-020-441-3 (1-16) x US-08-980-832-27 (1-11233)
QY 2 G|YrYrG|YArGApApApApApApApPProSeSerArg 14
Db 6700 GGAAGCGGTGCGCGATCCGCGCGATCGATGCGCCAGCGCG 6738

RESULT 11
US-09-272-032-7/C
Sequence 7, Application US/09272032A
Patent No. 6296852
GENERAL INFORMATION:
APPLICANT: Johnson, Michael A.
APPLICANT: Pridaux, Christopher T.
APPLICANT: McCoy, Richard J.
TITLE OF INVENTION: Recombinant Avian Adenovirus Vector
FILE REFERENCE: 48-95a
CURRENT APPLICATION NUMBER: US/09/272,032A
CURRENT FILING DATE: 1999-03-18
EARLIER APPLICATION NUMBER: US 08/448,617
EARLIER FILING DATE: 1995-09-08
EARLIER APPLICATION NUMBER: PCT/AU94/000189
EARLIER FILING DATE: 1994-04-14
EARLIER APPLICATION NUMBER: AU PL 8297
EARLIER FILING DATE: 1993-04-14
NUMBER OF SEQ ID NOS: 8
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 7
LENGTH: 8520
TYPE: DNA
ORGANISM: fowl adenovirus
US-09-272-032-7

Alignment Scores:
Pred. No.: 1.13e+03      Length: 8520
Score: 42.50      Matches: 9
Percent Similarity: 83.33%      Conservative: 1
Best Local Similarity: 75.00%      Mismatches: 1
Query Match: 47.22%      Indels: 1
DB: 4      Gaps: 1

US-10-020-441-3 (1-16) x US-09-272-032-7 (1-8520)
QY 1 Va|G|YrYr---G|YArGApApApApApApPProSeSerArg 11
Db 8010 GTAGGCTACGAGGATCGAGATGACGAGACCGCGAT 7975

RESULT 12
US-07-903-103-1
Sequence 1, Application US/07903103
Patent No. 5411860

```

```

GENERAL INFORMATION:
APPLICANT: VOGELSTEIN, BERT
APPLICANT: KINZLER, KENNETH
TITLE OF INVENTION: AMPLIFICATION OF HUMAN MDM2 GENE IN
CELL LINE: CaCo-2
NUMBER OF SEQUENCES: 4
CORRESPONDENCE ADDRESSES:
ADDRESSEE: BANNER, BIRCH, MCKIE AND BECKETT
STREET: 1001 G ST., N.W.
CITY: WASHINGTON
STATE: D.C.
COUNTRY: USA
ZIP: 20001-4597
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/903,103
FILING DATE: 19920623
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/867,840
FILING DATE: 07-APR-1992
ATTORNEY/AGENT INFORMATION:
NAME: KAGAN, SARAH A.
REGISTRATION NUMBER: 32,141
REFERENCE/DOCKET NUMBER: 01107.40148
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202-508-9100
TELEFAX: 202-508-9299
TELEX: 197430 BBMB UT
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 2372 base pairs
TYPE: NUCLEIC ACID
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: cDNA
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: Homo sapiens
CELL LINE: CaCo-2
POSITION IN GENOME:
MAP POSITION: 12q2-14
FEATURES:
NAME/KEY: CDS
LOCATION: 312..1784
US-07-903-103-1

Alignment Scores:
Pred. No.: 327 Length: 2372
Score: 42.00 Matches: 7
Percent Similarity: 90.00% Conservative: 2
Best Local Similarity: 70.00% Mismatches: 1
Query Match: 46.67% Indels: 0
DB: 1 Gaps: 0

US-10-020-441-3 (1-16) x US-07-903-103-1 (1-2372)
Oy 4 GIYARGASPAASnaSPATgAAPPProSer 13
|||||:::|||||::|
Db 84 GGCGGCGACCCCTCTGTACGAGATCTTGCT 113

RESULT 13
US-08-044-619A-1
; Sequence 1, Application US/08044619A
; Patent No. 5420263
GENERAL INFORMATION:
APPLICANT: THE JOHNS HOPKINS UNIVERSITY
APPLICANT: 720 RUTLAND AVENUE, BALTIMORE, MARYLAND 21205 USA

```

TITLE OF INVENTION: AMPLIFICATION OF HUMAN MD2 GENE IN  
NUMBER OF SEQUENCES: 4  
CORRESPONDENCE ADDRESSES:  
ADDRESSEE: BANNER, BIRCH, MCKIE AND BECKETT  
STREET: 1001 G ST., N.W.  
CITY: WASHINGTON  
STATE: D.C.  
COUNTRY: USA  
ZIP: 20001-4597  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/044,619A  
FILING DATE: 07-APR-1993  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/903,103  
FILING DATE: 23-JUN-1992  
APPLICATION NUMBER: US 07/867,840  
FILING DATE: 07-APR-1992  
ATTORNEY/AGENT INFORMATION:  
NAME: KAGAN, SARAH A.  
REGISTRATION NUMBER: 32,141  
REFERENCE/DOCKET NUMBER: 01107,40148  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 202-508-9100  
TELEFAX: 202-508-9299  
TELEX: 197430 BBMB UT  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 2372 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
MOLECULE TYPE: cDNA  
HYPOTHETICAL: NO  
ANTI-SENSE: NO  
ORIGINAL SOURCE:  
ORGANISM: Homo sapiens  
CELL LINE: CaCo-2  
POSITION IN GENOME:  
MAP POSITION: 12q12-14  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 312..1784  
US-08-044-619A-1

Alignment Scores:  
Pred. No.: 327 Length: 2372  
Score: 42.00 Matches: 7  
Percent Similarity: 90.00% Conservative: 2  
Best Local Similarity: 70.00% Mismatches: 1  
Query Match: 46.67% Indels: 0  
DB: 1 Gaps: 0

US-10-020-441-3 (1-16) x US-08-044-619A-1 (1-2372)

Qy 4 G1YARGASpASnASpARgASpProser 13  
Db 84 GGCCCGAGACCCTGTGACCGAGATCCTGCT 113

RESULT 14  
US-08-283-911-1  
Sequence 1, Application US/08283911  
Patent No. 5519118  
GENERAL INFORMATION:  
APPLICANT: VOGELSTEIN, BERT  
APPLICANT: KINZLER, KENNETH  
TITLE OF INVENTION: AMPLIFICATION OF HUMAN MD2 GENE IN

TITLE OF INVENTION: HUMAN TUMORS  
NUMBER OF SEQUENCES: 4  
CORRESPONDENCE ADDRESSES:  
ADDRESSEE: BANNER, BIRCH, MCKIE AND BECKETT  
STREET: 1001 G ST., N.W.  
CITY: WASHINGTON  
STATE: D.C.  
COUNTRY: USA  
ZIP: 20001-4597  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/283,911  
FILING DATE:  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/903,103  
FILING DATE: 23-JUN-1992  
APPLICATION NUMBER: US 07/867,840  
FILING DATE: 07-APR-1992  
ATTORNEY/AGENT INFORMATION:  
NAME: KAGAN, SARAH A.  
REGISTRATION NUMBER: 32,141  
REFERENCE/DOCKET NUMBER: 01107,40148  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 202-508-9100  
TELEFAX: 202-508-9299  
TELEX: 197430 BBMB UT  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 2372 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
MOLECULE TYPE: cDNA  
HYPOTHETICAL: NO  
ANTI-SENSE: NO  
ORIGINAL SOURCE:  
ORGANISM: Homo sapiens  
CELL LINE: CaCo-2  
POSITION IN GENOME:  
MAP POSITION: 12q12-14  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 312..1784  
US-08-283-911-1

Alignment Scores:  
Pred. No.: 327 Length: 2372  
Score: 42.00 Matches: 7  
Percent Similarity: 90.00% Conservative: 2  
Best Local Similarity: 70.00% Mismatches: 1  
Query Match: 46.67% Indels: 0  
DB: 1 Gaps: 0

US-10-020-441-3 (1-16) x US-08-283-911-1 (1-2372)

Qy 4 G1YARGASpASnASpARgASpProser 13  
Db 84 GGCCCGAGACCCTGTGACCGAGATCCTGCT 113

RESULT 15  
US-08-245-500A-2  
Sequence 2, Application US/08245500A  
Patent No. 5550023  
GENERAL INFORMATION:  
APPLICANT: BURRELL, MARILEE  
APPLICANT: HILL, DAVID E.  
APPLICANT: KINZLER, KENNETH W.  
APPLICANT: VOGELSTEIN, BERT



COUNTRY: USA  
ZIP: 20001  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/390,479A  
FILING DATE: 02-FEB-1995  
CLASSIFICATION: 530  
ATTORNEY/AGENT INFORMATION:  
NAME: KAGAN, SARAH A.  
REGISTRATION NUMBER: 32,141  
REFERENCE/DOCKET NUMBER: 01107.48992  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 202-508-9100  
TELEFAX: 202-508-9299  
TELEX: 197430 BBWM UT  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 2372 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
MOLECULE TYPE: cDNA  
HYPOTHETICAL: NO  
ANTI-SENSE: NO  
ORIGINAL SOURCE:  
ORGANISM: Homo sapiens  
CELL LINE: CaCo-2  
POSITION IN GENOME:  
MAP POSITION: 12q12-14  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 312..1784  
US-08-390-479A-2

Alignment Scores:	
Pred. No.:	337
Score:	42.00
Percent Similarity:	90.00%
Best Local Similarity:	70.00%
Query Matchn:	46.67%
DB:	1
	Gaps: 0
US-10-020-441-3 (1-16) x US-08-350-479A-2 (1-2372)	2372

US-10-020-441-3 (1-16) x US-08-390-479A-2 (1-2372)

QY 4 G|yArqAspAsnAspArgAspProSer 13  
          :::|||||||  
Db 84 GCCCGGACCCCTCTTACCAGATCCTGCT 113

```

RESULT 18
US-08-557-393-2
; Sequence 2, Application US/08557393
; Patent No. 5702903
; GENERAL INFORMATION:
; APPLICANT: BURRELL, MARILEE
; APPLICANT: HILL, DAVID E.
; APPLICANT: KENZLER, KENNETH W.
; APPLICANT: VOGELSTEIN, BERT
; TITLE OF INVENTION: AMPLIFICATION OF HUMAN MDM2 GENE IN
; TISSUE OF INVENTION: HUMAN TUMORS
; NUMBER OF SEQUENCES: 5
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: BANNER, BIRCH, MCKIE AND BECKETT
; STREET: 1001 G STREET, N.W.
; CITY: WASHINGTON
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20001
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
;

```

```

COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/557,393
FILING DATE: 13-NOV-1995
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/245,500
FILING DATE: 18-MAY-1994
ATTORNEY/AGENT INFORMATION:
NAME: KAGAN, SARAH A.
REGISTRATION NUMBER: 32,141
REFERENCE/DOCKET NUMBER: 01107.42798
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202-508-9100
TELEFAX: 202-508-9299
TELEX: 197430 BMB UT
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 2372 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: cDNA
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: Homo sapiens
CELL LINE: CaCo-2
POSITION IN GENOME:
MAP POSITION: 12q12-14
FEATURE:
NAME/KEY: CDS
LOCATION: 312..1784

```

Alignment Scores:	
Pred. No.:	327
Score:	42.00
Percent Similarity:	90.00
Best Local Similarity:	70.00%
Query Match:	46.67%
DB:	1
Length:	2372
Matches:	777
Conservative:	2
Mismatches:	1
Indels:	0
Gaps:	0

US-10-020-441-3 (1-16) x US-08-557-393-2 (1-2372)

Dy 4 GLYARGASPASPA~~SL~~PARGASPProSer 13  
| | | | | : | | | | | : : :  
Db 84 GGC CGACCCCTGTGACCGAGATCCTGCT 113

```

US-08-390-516C-2
; Sequence 2, Application US/08390516C
; Patent No. 5708136
;
; GENERAL INFORMATION:
; APPLICANT: BURRELL, MARILEE
; APPLICANT: HILL, DAVID E.
; APPLICANT: KINZLER, KENNETH W.
; APPLICANT: VOGELSTEIN, BERT
; TITLE OF INVENTION: AMPLIFICATION OF HUMAN MDM2 GENE IN
; TITLE OF INVENTION: HUMAN TUMORS
; NUMBER OF SEQUENCES: 9
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: BANNER, BIRCH, MCKIE AND BECKETT
; STREET: 1001 G STREET, N.W.
; CITY: WASHINGTON
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20001
;
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
;

```

```

; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/390,516C
; FILING DATE: 07-APR-1993
; CLASSIFICATION: 530
; ATTORNEY/AGENT INFORMATION:
; NAME: KAGAN, SARAH A.
; REGISTRATION NUMBER: 32,141
; REFERENCE/DOCKET NUMBER: 01107.42798
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-508-9100
; TELEFAX: 202-508-9299
; TELEX: 197430 BEMB UT
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2372 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; HYPOTHEICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: Homo sapiens
; CELL LINE: Caco-2
; POSITION IN GENOME:
; MAP POSITION: 12q12-14
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 312..1784
; US-08-390-516C-2

Alignment Scores:
Pred. No.: 327 Length: 2372
Score: 42.00 Matches: 7
Percent Similarity: 90.00% Conservative: 2
Best Local Similarity: 70.00% Mismatches: 1
Query Match: 46.67% Indels: 0
DB: 1 Gaps: 0

US-10-020-441-3 (1-16) x US-08-390-516C-2 (1-2372)

Cy 4 G1YArGaPaPaPaPaPaPaPaPaPaProSer 13
Db 84 GGCCGCGACCCCTCTGACCGAGATCTGCT 113

RESULT 20
; US-08-390-517A-2
; Sequence 2, Application US/08390517A
; Patent No. 573638
; GENERAL INFORMATION:
; APPLICANT: BURRELL, MARILEE
; APPLICANT: HILL, DAVID E.
; APPLICANT: KINZLER, KENNETH W.
; APPLICANT: VOGELSTEIN, BERT
; TITLE OF INVENTION: AMPLIFICATION OF HUMAN MDM2 GENE IN
; TITLE OF INVENTION: HUMAN TUMORS
; NUMBER OF SEQUENCES: 5
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: BANNER, BIRCH, MCKIE AND BECKETT
; STREET: 1001 G STREET, N.W.
; CITY: WASHINGTON
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20001
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/390,517A
```

```

; FILING DATE: 07-APR-1993
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: KAGAN, SARAH A.
; REGISTRATION NUMBER: 32,141
; REFERENCE/DOCKET NUMBER: 01107.42798
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-508-9100
; TELEFAX: 202-508-9299
; TELEX: 197430 BEMB UT
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2372 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; HYPOTHEICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: Homo sapiens
; CELL LINE: Caco-2
; POSITION IN GENOME:
; MAP POSITION: 12q12-14
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 312..1784
; US-08-390-517A-2

Alignment Scores:
Pred. No.: 327 Length: 2372
Score: 42.00 Matches: 7
Percent Similarity: 90.00% Conservative: 2
Best Local Similarity: 70.00% Mismatches: 1
Query Match: 46.67% Indels: 0
DB: 1 Gaps: 0

US-10-020-441-3 (1-16) x US-08-390-517A-2 (1-2372)

Cy 4 G1YArGaPaPaPaPaPaPaPaPaPaProSer 13
Db 84 GGCCGCGACCCCTCTGACCGAGATCTGCT 113

RESULT 21
; US-08-390-515A-2
; Sequence 2, Application US/08390515A
; Patent No. 5756455
; GENERAL INFORMATION:
; APPLICANT: BURRELL, MARILEE
; APPLICANT: HILL, DAVID E.
; APPLICANT: KINZLER, KENNETH W.
; APPLICANT: VOGELSTEIN, BERT
; TITLE OF INVENTION: AMPLIFICATION OF HUMAN MDM2 GENE IN
; TITLE OF INVENTION: HUMAN TUMORS
; NUMBER OF SEQUENCES: 9
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: BANNER, BIRCH, MCKIE AND BECKETT
; STREET: 1001 G STREET, N.W.
; CITY: WASHINGTON
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20001
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/390,515A
; FILING DATE: 07-APR-1993
; CLASSIFICATION: 514
; ATTORNEY/AGENT INFORMATION:
; NAME: KAGAN, SARAH A.
```

REGISTRATION NUMBER: 32,141  
REFERENCE/DOCKET NUMBER: 01107.42798  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 202-508-9100  
TELEFAX: 202-508-9299  
TELEX: 197430 BMB UT  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 2372 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
MOLECULE TYPE: cDNA  
HYPOTHETICAL: NO  
ANTI-SENSE: NO  
ORIGINAL SOURCE:  
ORGANISM: Homo sapiens  
CELL LINE: Caco-2  
POSITION IN GENOME:  
MAP POSITION: 12q12-14  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 312..1784  
US-08-390-515A-2

Alignment Scores:  
Pred. No.: 327 Length: 2372  
Score: 42.00 Matches: 7  
Percent Similarity: 90.00% Conservative: 2  
Best Local Similarity: 70.00% Mismatches: 1  
Query Match: 46.67% Indels: 0  
DB: 1 Gaps: 0

US-10-020-441-3 (1-16) x US-08-390-515A-2 (1-2372)

Qy 4 G1YArGAsPAsPAsnAsPArGAsPProSer 13  
Db 84 GGCCCGAGACCCCTGTGACCGAGATCCTGCT 113

RESULT 22  
US-08-801-718-2  
Sequence 2, Application US/08801718  
Patent No. 5858976  
GENERAL INFORMATION:  
APPLICANT: BURRELL, MARILEE  
APPLICANT: HILL, DAVID E.  
APPLICANT: KINZLER, KENNETH W.  
TITLE OF INVENTION: AMPLIFICATION OF HUMAN MDM2 GENE IN  
TITLE OF INVENTION: HUMAN TUMORS  
NUMBER OF SEQUENCES: 9  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: BANNER, BIRCH, MCKIE AND BECKETT  
STREET: 1001 G STREET, N.W.  
CITY: WASHINGTON  
STATE: D.C.  
COUNTRY: USA  
ZIP: 20001  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/801,718  
FILING DATE: 14-FEB-1997  
CLASSIFICATION: 514  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/390,515  
FILING DATE: 07-APR-1993  
ATTORNEY/AGENT INFORMATION:  
NAME: KAGAN, SARAH A.  
REGISTRATION NUMBER: 32,141

REFERENCE/DOCKET NUMBER: 01107.42798  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 202-508-9100  
TELEFAX: 202-508-9299  
TELEX: 197430 BMB UT  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 2372 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
MOLECULE TYPE: cDNA  
HYPOTHETICAL: NO  
ANTI-SENSE: NO  
ORIGINAL SOURCE:  
ORGANISM: Homo sapiens  
CELL LINE: Caco-2  
POSITION IN GENOME:  
MAP POSITION: 12q12-14  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 312..1784  
US-08-801-718-2

Alignment Scores:  
Pred. No.: 327 Length: 2372  
Score: 42.00 Matches: 7  
Percent Similarity: 90.00% Conservative: 2  
Best Local Similarity: 70.00% Mismatches: 1  
Query Match: 46.67% Indels: 0  
DB: 2 Gaps: 0

US-10-020-441-3 (1-16) x US-08-801-718-2 (1-2372)

Qy 4 G1YArGAsPAsPAsnAsPArGAsPProSer 13  
Db 84 GGCCCGAGACCCCTGTGACCGAGATCCTGCT 113

RESULT 23  
US-09-073-567-1  
Sequence 1, Application US/09073567  
Patent No. 6013786  
GENERAL INFORMATION:  
APPLICANT: JIANDONG CHEN  
APPLICANT: SUDHIR AGRAWAL  
APPLICANT: RUIWEN ZHANG  
TITLE OF INVENTION: MDM2-SPECIFIC ANTISENSE OLIGONUCLEOTIDES  
NUMBER OF SEQUENCES: 49  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: McDONNELL BOEHLEN HULBERT & BERGHOFF  
STREET: 300 South Wacker Drive, 32nd Floor  
CITY: Chicago  
STATE: IL  
COUNTRY: United States of America  
ZIP: 60606  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Microsoft Word 97  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/073,567  
FILING DATE:  
CLASSIFICATION:  
ATTORNEY/AGENT INFORMATION:  
NAME: Greenfield, Michael S.  
REGISTRATION NUMBER: 37,147  
REFERENCE/DOCKET NUMBER: 98,057-A  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (312) 913-0001  
TELEFAX: (312) 913-0002  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:



LENGTH: 2372 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: both  
TOPOLOGY: linear  
MOLECULE TYPE: hmdm2 DNA  
HYPOTHETICAL: NO  
ANTI-SENSE: NO  
US-09-073-567-1

Alignment Scores:  
Pred. No.: 327 Length: 2372  
Score: 42.00 Matches: 7  
Percent Similarity: 90.00% Conservative: 2  
Best Local Similarity: 70.00% Mismatches: 1  
Query Match: 46.67% Indels: 0  
DB: 3 Gaps: 0

US-10-020-441-3 (1-16) x US-09-073-567-1 (1-2372)

Qy 4 G1YARGAPAPAsnAspArgAspProSer 13  
Db 84 GGGCCGACCCCTCTGACCGAGATCTGCT 113

RESULT 24  
US-09-280-805-1  
Sequence 1, Application US/09280805  
Patent No. 6184212  
GENERAL INFORMATION:  
APPLICANT: Loren J. Miraglia, Pamela Nero, Mark J.  
APPLICANT: Graham, Brett P. Monia  
TITLE OF INVENTION: ANTISENSE MODULATION OF HUMAN MDM2  
NUMBER OF SEQUENCES: 271  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Law Offices of Jane Massey Licata  
STREET: 66 East Main Street  
CITY: Marlton  
STATE: NJ  
COUNTRY: U.S.A.  
ZIP: 08053  
COMPUTER READABLE FORM:  
MEDIUM TYPE: DISKETTE, 3.5 INCH, 1.44 MB STORAGE  
COMPUTER: IBM PC  
OPERATING SYSTEM: WINDOWS 95  
SOFTWARE: WORDPERFECT 6.0  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/280,805  
FILING DATE: herewith  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 09/048,810  
FILING DATE: March 26, 1998  
ATTORNEY/AGENT INFORMATION:  
NAME: Licata, Jane Massey  
REGISTRATION NUMBER: 32,257  
REFERENCE/DOCKET NUMBER: ISPH-0346  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 609-810-1515  
TELEFAX: 609-810-1454  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 2372 base pairs  
TYPE: Nucleic Acid  
STRANDEDNESS: Single  
TOPOLOGY: Unknown  
ANTI-SENSE: NO  
PUBLICATION INFORMATION:  
AUTHORS: Oliner,J.D.  
AUTHORS: Kinzler,K.W.  
AUTHORS: Meltzer,P.S.  
AUTHORS: George,D.L.  
AUTHORS: Vogelstein,B.  
TITLE: Amplification of a gene encoding a

TITLE: p53-associated protein in human sarcomas  
JOURNAL: Nature  
VOLUME: 358  
ISSUE: 6381  
PAGES: 80-83  
DATE: 02-JUL-1992  
US-09-280-805-1

Alignment Scores:  
Pred. No.: 327 Length: 2372  
Score: 42.00 Matches: 7  
Percent Similarity: 90.00% Conservative: 2  
Best Local Similarity: 70.00% Mismatches: 1  
Query Match: 46.67% Indels: 0  
DB: 4 Gaps: 0

US-10-020-441-3 (1-16) x US-09-280-805-1 (1-2372)

Qy 4 G1YARGAPAPAsnAspArgAspProSer 13  
Db 84 GGGCCGACCCCTCTGACCGAGATCTGCT 113

RESULT 25  
US-09-048-810-1  
Sequence 1, Application US/09048810  
Patent No. 6238921  
GENERAL INFORMATION:  
APPLICANT: Loren J. Miraglia, Pamela Nero, Mark J.  
APPLICANT: Graham, Brett P. Monia  
TITLE OF INVENTION: ANTISENSE OLIGONUCLEOTIDE  
NUMBER OF SEQUENCES: 32  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Law Offices of Jane Massey Licata  
STREET: 66 East Main Street  
CITY: Marlton  
STATE: NJ  
COUNTRY: U.S.A.  
ZIP: 08053  
COMPUTER READABLE FORM:  
MEDIUM TYPE: DISKETTE, 3.5 INCH, 1.44 MB STORAGE  
COMPUTER: IBM 486  
OPERATING SYSTEM: WINDOWS FOR WORKGROUPS  
SOFTWARE: WORDPERFECT 5.1  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/048,810  
FILING DATE: herewith  
CLASSIFICATION:  
ATTORNEY/AGENT INFORMATION:  
NAME: Licata, Jane Massey  
REGISTRATION NUMBER: 32,257  
REFERENCE/DOCKET NUMBER: ISPH-0302  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 609-779-2400  
TELEFAX: 609-810-1454  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 2372 base pairs  
TYPE: Nucleic Acid  
STRANDEDNESS: Single  
TOPOLOGY: Unknown  
ANTI-SENSE: NO  
PUBLICATION INFORMATION:  
AUTHORS: Oliner,J.D.  
AUTHORS: Kinzler,K.W.  
AUTHORS: Meltzer,P.S.  
AUTHORS: George,D.L.  
AUTHORS: Vogelstein,B.  
TITLE: Amplification of a gene encoding a  
TITLE: p53-associated protein in human sarcomas  
JOURNAL: Nature  
VOLUME: 358  
ISSUE: 6381

PAGES: 80-83  
DATE: 02-JUL-1992  
US-09-048-810-1

Alignment Scores:  
Pred. No.: 327 Length: 2372  
Score: 42.00 Matches: 7  
Percent Similarity: 90.00% Conservative: 2  
Best Local Similarity: 70.00% Mismatches: 1  
Query Match: 46.67% Indels: 0  
DB: 4 Gaps: 0

US-10-020-441-3 (1-16) x US-09-048-810-1 (1-2372)

OY 4 G1YAGAspAspAsnAspArgAspProser 13  
DB 84 GGCCCGAGACCCCTGTGACCGAGATCTGCT 113

RESULT 26  
US-09-170-159A-2  
Sequence 2, Application US/09170159A  
Patent No. 639755  
GENERAL INFORMATION:  
APPLICANT: BURRELL, MARILEE  
HILL, DAVID E.  
KINZLER, KENNETH W.  
VOGELSTEIN, BERT  
TITLE OF INVENTION: AMPLIFICATION OF HUMAN MDM2 GENE IN  
VOGELSTEIN, BERT  
NUMBER OF SEQUENCES: 5  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: BANNER, BIRCH, MCKIE AND BECKETT  
STREET: 1001 G STREET, N.W.  
CITY: WASHINGTON  
STATE: D.C.  
COUNTRY: USA  
ZIP: 20001  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/170,159A  
FILING DATE: 13-Oct-1998  
CLASSIFICATION: <Unknown>  
ATTORNEY/AGENT INFORMATION:  
NAME: KAGAN, SARAH A.  
REGISTRATION NUMBER: 32,141  
REFERENCE/DOCKET NUMBER: 01107,42798  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 202-508-9100  
TELEFAX: 202-508-9299  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 2372 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
MOLECULE TYPE: cDNA  
HYPOTHETICAL: NO  
ANTI-SENSE: NO  
ORIGINAL SOURCE:  
ORGANISM: Homo sapiens  
CELL LINE: CaCo-2  
POSITION IN GENOME:  
MAP POSITION: 12q12-14  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 312..1784  
SEQUENCE DESCRIPTION: SEQ ID NO: 2:  
US-09-170-159A-2

Alignment Scores:  
Pred. No.: 327 Length: 2372  
Score: 42.00 Matches: 7  
Percent Similarity: 90.00% Conservative: 2  
Best Local Similarity: 70.00% Mismatches: 1  
Query Match: 46.67% Indels: 0  
DB: 4 Gaps: 0

US-10-020-441-3 (1-16) x US-09-170-159A-2 (1-2372)

OY 4 G1YAGAspAspAsnAspArgAspProser 13  
DB 84 GGCCCGAGACCCCTGTGACCGAGATCTGCT 113

RESULT 27  
US-09-480-718-43  
Sequence 43, Application US/09480718  
Patent No. 640762  
GENERAL INFORMATION:  
APPLICANT: Sherr, Charles J  
APPLICANT: Quelle, Dawn E.  
APPLICANT: Weber, Jason D.  
APPLICANT: Roussel, Martine F.  
APPLICANT: Frederique, Zindy  
TITLE OF INVENTION: ARF-19, A NOVEL REGULATOR OF THE MAMMALIAN CELL CYCLE  
FILE REFERENCE: 1340-1-023 CIP 1  
CURRENT APPLICATION NUMBER: US/09/480,718  
EARLIER FILING DATE: 09/129,855  
NUMBER OF SEQ ID NOS: 48  
SOFTWARE: Patentin Ver. 2.0  
SEQ ID NO 43  
LENGTH: 2372  
TYPE: DNA  
ORGANISM: Homo sapiens  
US-09-480-718-43

Alignment Scores:  
Pred. No.: 327 Length: 2372  
Score: 42.00 Matches: 7  
Percent Similarity: 90.00% Conservative: 2  
Best Local Similarity: 70.00% Mismatches: 1  
Query Match: 46.67% Indels: 0  
DB: 4 Gaps: 0

US-10-020-441-3 (1-16) x US-09-480-718-43 (1-2372)

OY 4 G1YAGAspAspAsnAspArgAspProser 13  
DB 84 GGCCCGAGACCCCTGTGACCGAGATCTGCT 113

RESULT 28  
US-09-106-486-3/c  
Sequence 3, Application US/09106486  
Patent No. 6043221  
GENERAL INFORMATION:  
APPLICANT: Magal, Elia  
APPLICANT: Delaney, John W.  
TITLE OF INVENTION: METHOD FOR PREVENTING AND TREATING  
TITLE OF INVENTION: HEARING LOSS USING A NEURTURIN PROTEIN PRODUCT  
NUMBER OF SEQUENCES: 5  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Amgen Inc.  
STREET: One Amgen Center Drive  
CITY: Thousand Oaks  
STATE: California  
COUNTRY: USA  
ZIP: 91320-1789  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/106,486  
FILING DATE: 29-JUN-1998  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 60/054184  
FILING DATE: 30-JUL-1997  
ATTORNEY/AGENT INFORMATION:  
NAME: Curry, Daniel R.  
REGISTRATION NUMBER: 32,727  
REFERENCE/DOCKET NUMBER: A-444  
INFORMATION FOR SEQ ID NO: 3:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 312 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 1..309  
US-09-106-486-3

Alignment Scores:  
Pred. No.: 49.4 Length: 312  
Score: 41.00 Matches: 8  
Percent Similarity: 60.00% Conservative: 1  
Best Local Similarity: 53.33% Mismatches: 6  
Query Match: 45.56% Indels: 0  
DB: 3 Gaps: 0

US-10-020-441-3 (1-16) x US-09-106-486-3 (1-312)  
Cy 2 G1YTYG1YArgAspAsnAspAspArgAspProSerArgLYAsn 16  
Db 178 GGCGACGGCGCTGACGACGACGACGACGACGACGTCGTAACAC 134

RESULT 29  
US-09-377-446-3/C  
Sequence 3, Application US/09377446  
Patent No. 6274554  
GENERAL INFORMATION:  
APPLICANT: Magal, Ella  
TITLE OF INVENTION: METHOD FOR PREVENTING AND TREATING  
HEARING LOSS USING A NEURTURIN PROTEIN PRODUCT  
NUMBER OF SEQUENCES: 5  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Amgen Inc.  
STREET: One Angen Center Drive  
CITY: Thousand Oaks  
STATE: California  
COUNTRY: USA  
ZIP: 91320-1789  
COMPUTER READABLE FORM:  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/377,446  
FILING DATE: 19-Aug-1999  
CLASSIFICATION: <unknown>  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 09/106,486  
FILING DATE: <unknown>  
ATTORNEY/AGENT INFORMATION:  
NAME: Curry, Daniel R.  
REGISTRATION NUMBER: 32,727  
REFERENCE/DOCKET NUMBER: A-444  
INFORMATION FOR SEQ ID NO: 3:

SEQUENCE CHARACTERISTICS:  
LENGTH: 312 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 1..309  
SEQUENCE DESCRIPTION: SEQ ID NO: 3:  
US-09-377-446-3

Alignment Scores:  
Pred. No.: 49.4 Length: 312  
Score: 41.00 Matches: 8  
Percent Similarity: 60.00% Conservative: 1  
Best Local Similarity: 53.33% Mismatches: 6  
Query Match: 45.56% Indels: 0  
DB: 4 Gaps: 0

US-10-020-441-3 (1-16) x US-09-377-446-3 (1-312)  
Cy 2 G1YTYG1YArgAspAsnAspAspArgAspProSerArgLYAsn 16  
Db 178 GGCGACGGCGCTGACGACGACGACGACGACGACGTCGTAACAC 134

RESULT 30  
US-09-512-342-19  
Sequence 19, Application US/09512342  
Patent No. 638068  
GENERAL INFORMATION:  
APPLICANT: SATOH, SHINOBU  
TITLE OF INVENTION: METHOD FOR PRODUCING FOREIGN POLYPEPTIDE IN PLANT  
FILE REFERENCE: 081356/0142  
CURRENT APPLICATION NUMBER: US/09/512,342  
CURRENT FILING DATE: 2000-02-24  
NUMBER OF SEQ ID NOS: 38  
SOFTWARE: Patentin Ver. 2.1  
SEQ ID NO 19  
LENGTH: 504  
TYPE: DNA  
ORGANISM: Cucumis sativus  
US-09-512-342-19

Alignment Scores:  
Pred. No.: 84.9 Length: 504  
Score: 41.00 Matches: 7  
Percent Similarity: 90.00% Conservative: 2  
Best Local Similarity: 70.00% Mismatches: 1  
Query Match: 45.56% Indels: 0  
DB: 4 Gaps: 0

US-10-020-441-3 (1-16) x US-09-512-342-19 (1-504)  
Cy 1 ValG1YTYG1YArgAspAsnAspArg 10  
Db 157 GTAGGTACGACGACGACCATCATGATCAA 186

RESULT 31  
US-09-221-017B-63/C  
Sequence 63, Application US/09221017B  
Patent No. 6444799  
GENERAL INFORMATION:  
APPLICANT: Ross, Bruce C.  
TITLE OF INVENTION: P. GINGIVALIS NUCLEOTIDES AND USES THEREOF  
NUMBER OF SEQUENCES: 1120  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: MORRISON & FOERSTER  
STREET: 755 PAGE MILL ROAD  
CITY: Palo Alto  
STATE: CA

```

; COUNTRY: USA
; ZIP: 94304-1018
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: Windows
; SOFTWARE: FastSeq for Windows Version 2.0b
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/221.017B
; FILING DATE: 23-DEC-1998
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PP1182
; FILING DATE: 31-DEC-1997
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PP1546
; FILING DATE: 30-JAN-1998
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PP2911
; FILING DATE: 09-APR-1998
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/AU98/01023
; FILING DATE: 10-DEC-1998
; ATTORNEY/AGENT INFORMATION:
; NAME: Monroy, Gladys H
; REGISTRATION NUMBER: 32,430
; REFERENCE/DOCKET NUMBER: 27340-20021.00
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 650-813-5600
; TELEFAX: 650-494-0792
; TELEX: 706141
; INFORMATION FOR SEQ ID NO: 63:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 670 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: circular
; MOLECULE TYPE: DNA (genomic)
; HYPOTHEetical: NO
; ANTI-SENSE: UNKNOWN
; ORIGINAL SOURCE:
; ORGANISM: FORYPHYROMONAS GINGIVALIS
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: 1...670
; US-09-221-017B-63
;
; Alignment Scores:
; Pred. No.: 117 Length: 670
; Score: 41.00 Matches: 10
; Percent Similarity: 55.00% Conservative: 1
; Best Local Similarity: 50.00% Mismatches: 3
; Query Match: 45.56% Indels: 6
; DB: 4 Gaps: 1
;
US-10-020-441-3 (1-16) x US-09-221-017B-63 (1-670)
Qy 1 ValG|YTYG|Y-----ArgAspAspAsnAspArgAspProSerArg 14
Db 345 GTCGGATACGGTTTACTGCTTCTCGCACATCGGCCAGACGAGATCGGAGAACAGTCTG 286
;
RESULT 32
US-09-048-889-9/c
; Sequence 9, Application US/09048889
; Patent No. 6117989
; GENERAL INFORMATION:
; APPLICANT: Bandman, Olga
; APPLICANT: Hillman, Jennifer L.
; APPLICANT: Corley, Neil C.
; APPLICANT: Guegler, Karl J.
; APPLICANT: Lal, Preeti
; APPLICANT: Patterson, Chandra
; TITLE OF INVENTION: HUMAN CALCIUM-BINDING PROTEINS
;

```

```

; NUMBER OF SEQUENCES: 11
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Incyte Pharmaceuticals, Inc.
; STREET: 3174 Porter Drive
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/048,889
; FILING DATE: Herewith
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Ceirone, Michael C
; REGISTRATION NUMBER: 39,132
; REFERENCE/DOCKET NUMBER: PF-0493 US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 650-855-0555
; TELEFAX: 650-845-4166
; TELEX:
; INFORMATION FOR SEQ ID NO: 9:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1496 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; LIBRARY: ISLNOT01
; CLONE: 2379155
; US-09-048-889-9
;
; Alignment Scores:
; Pred. No.: 289 Length: 1496
; Score: 41.00 Matches: 7
; Percent Similarity: 78.57% Conservative: 4
; Best Local Similarity: 50.00% Mismatches: 3
; Query Match: 45.56% Indels: 0
; DB: 3 Gaps: 0
;
US-10-020-441-3 (1-16) x US-09-048-889-9 (1-1496)
Qy 3 TYRG|YArgAspAspAsnAspArgAspProSerArgLysAsn 16
Db 178 CATGGTGTTCAGCAGAACGACGACCGGCGCGAAGAGAC 137
;
RESULT 33
US-08-796-899-22
; Sequence 22, Application US/08796899
; Patent No. 6160202
; GENERAL INFORMATION:
; APPLICANT: BUSTOS, Mauricio M
; APPLICANT: CHERN, Maw-Sheng
; TITLE OF INVENTION: MODIFICATION OF SEED CROPS WITH
; NUMBER OF SEQUENCES: 32
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Burns, Doane, Swecker & Mathis
; STREET: P.O. Box 1404
; CITY: Alexandria
; STATE: Virginia
; COUNTRY: United States
; ZIP: 22313-1404
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
;

```

```

; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/796,899
; FILING DATE: 06-FEB-1997
; CLASSIFICATION: 800
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/319,544
; FILING DATE: 07-OCT-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Meuth, Donna M
; REGISTRATION NUMBER: 36,607
; REFERENCE/DOCKET NUMBER: 028754-005
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 836-6620
; TELEFAX: (703) 836-2021
; INFORMATION FOR SEQ ID NO: 22:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1546 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; US-08-796-899-22

Alignment Scores:
Pred. No.: 300 Length: 1546
Score: 41.00 Matches: 8
Percent Similarity: 64.29% Conservative: 1
Best Local Similarity: 57.14% Mismatches: 5
Query Match: 45.56% Indels: 0
DB: 4 Gaps: 0

US-10-020-441-3 (1-16) x US-08-796-899-22 (1-1546)
QY 2 G1YTYG1YARGASPASnAspAgtgAProSeArGlyS 15
Db 813 GGAAGGGAAGTGAAGATCGAAGATCGATCCCAAAAAA 854

RESULT 34
US-09-434-288-5/c
; Sequence 5, Application US/09434288
; Patent No. 6303767
; GENERAL INFORMATION:
; APPLICANT: Betlach C, Melanie
; TITLE OF INVENTION: POLYKETIDE SYNTHASE ENZYMES AND RECOMBINANT DNA
; FILE REFERENCE: 30062-20030.00
; CURRENT APPLICATION NUMBER: US/09/434,288
; CURRENT FILING DATE: 1999-11-05
; PRIOR APPLICATION NUMBER: 60/107,093
; PRIOR FILING DATE: 1998-11-05
; NUMBER OF SEQ ID NOS: 12
; SOFTWARE: Patentin Ver. 2.1
; SEQ ID NO 5
; LENGTH: 1881
; TYPE: DNA
; ORGANISM: Streptomyces narbonensis
; US-09-434-288-5

Alignment Scores:
Pred. No.: 374 Length: 1881
Score: 41.00 Matches: 7
Percent Similarity: 75.00% Conservative: 2
Best Local Similarity: 58.33% Mismatches: 3
Query Match: 45.56% Indels: 0
DB: 4 Gaps: 0

US-10-020-441-3 (1-16) x US-09-434-288-5 (1-1881)
QY 4 G1YARGASPASnAspAgtgAProSeArGlyS 15
|||||
::: |||||
::: |||||
```

```

Db 1847 GGACGAGACCTGACAAAGAGACCCGCGCTCGA 1812
RESULT 35
US-08-336-308A-9/c
; Sequence 9, Application US/08336308A
; Patent No. 6017532
; GENERAL INFORMATION:
; APPLICANT: Travis, James
; APPLICANT: Potempa, Jan S.
; APPLICANT: Barr, Philip J.
; APPLICANT: Pavloff, Nadine
; TITLE OF INVENTION: Porphyromonas gingivalis
; TITLE OF INVENTION: Arginine-specific Proteinase Coding Sequences
; NUMBER OF SEQUENCES: 16
; CORRESPONDENCE ADDRESSES:
; ADDRESSEE: Greenlee, Winner and Sullivan, P.C.
; STREET: 5370 Mammoth Circle, Suite 201
; CITY: Boulder
; STATE: Colorado
; COUNTRY: US
; ZIP: 80303
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/336,308A
; FILING DATE: 08-NOV-1994
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/119,361
; FILING DATE: 10-SEP-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/265,441
; FILING DATE: 24-JUN-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Feiber, Donna M.
; REGISTRATION NUMBER: 33,878
; REFERENCE/DOCKET NUMBER: 21-93C
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (303) 499-8080
; TELEFAX: (303) 499-8089
; INFORMATION FOR SEQ ID NO: 9:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 7266 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHEICAL: NO
; ANTI-SENSE: NO
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 949..6063
; US-08-336-308A-9

Alignment Scores:
Pred. No.: 1,72e+03 Length: 7266
Score: 41.00 Matches: 7
Percent Similarity: 72.73% Conservative: 1
Best Local Similarity: 63.64% Mismatches: 3
Query Match: 45.56% Indels: 0
DB: 3 Gaps: 0

US-10-020-441-3 (1-16) x US-08-336-308A-9 (1-7266)
QY 3 TyrG1YARGASPASnAspAgtgAProSeAr 13
|||||
|||||
::: |||||
::: |||||

Db 6380 TATGACGGAAGTGAAGACCCGCGCTCTGCG 6348

RESULT 36
US-08-822-324-5/c
```

; Sequence 5, Application US/08822324  
; Patent No. 6129917  
; GENERAL INFORMATION:  
; APPLICANT: Potempa, Jan S.  
; APPLICANT: Travis, James  
; APPLICANT: Genco, Caroline A.  
; TITLE OF INVENTION: IMMUNOGENIC COMPOSITIONS COMPRISING  
; TITLE OF INVENTION: PORPHYROMONAS GINGIVALIS PROTEINS AND/OR PEPTIDES AND  
; TITLE OF INVENTION: METHODS  
; NUMBER OF SEQUENCES: 24  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Greenlee, Winner and Sullivan, P.C.  
; STREET: 5370 Manhattan Circle, Suite 201  
; CITY: Boulder  
; STATE: CO  
; COUNTRY: US  
; ZIP: 80303  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentin Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/822,324  
; FILING DATE: 21-MAR-1997  
; CLASSIFICATION: 536  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 60/013,945  
; FILING DATE: 22-MAR-1996  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Feiber, Donna M.  
; REGISTRATION NUMBER: 33,878  
; REFERENCE/DOCKET NUMBER: 103-95 WO  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (303) 488-8080  
; TELEFAX: (303) 499-8089  
; INFORMATION FOR SEQ ID NO: 5:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 7266 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: double  
; TOPOLOGY: linear  
; MOLECULE TYPE: DNA (genomic)  
; HYPOTHETICAL: NO  
; ORIGINAL SOURCE:  
; ORGANISM: Porphyromonas gingivalis  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: 949..6063  
; US-08-822-324-5  
  
Alignment Scores:  
Pred. No.: 1,72e+03 Length: 7266  
Score: 41.00 Matches: 7  
Percent Similarity: 72.73% Conservative: 1  
Best Local Similarity: 63.64% Mismatches: 3  
Query Match: 45.56% Indels: 0  
DB: 3 Gaps: 0  
  
US-10-020-441-3 (1-16) x US-08-822-324-5 (1-7266)  
Qy 3 TyrglyArgAspAspAsnAspArgAspProser 13  
Db 6380 TATGACGGGACTCGAACCCGCGACCCCTGCG 6348  
  
RESULT 37  
US-09-490-931-9/c  
; Sequence 9, Application US/09490931  
; Patent No. 6274718  
; GENERAL INFORMATION:  
; APPLICANT: Travis, James  
; APPLICANT: Potempa, Jan S.  
; APPLICANT: Barr, Philip J.

; APPLICANT: Pavloff, Nadine  
; TITLE OF INVENTION: Porphyromonas gingivalis  
; TITLE OF INVENTION: Arginine-specific Proteinase Coding Sequences  
; NUMBER OF SEQUENCES: 16  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Greenlee, Winner and Sullivan, P.C.  
; STREET: 5370 Manhattan Circle, Suite 201  
; CITY: Boulder  
; STATE: Colorado  
; COUNTRY: US  
; ZIP: 80303  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentin Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/490,931  
; FILING DATE:  
; CLASSIFICATION:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 08/336,308  
; FILING DATE:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/265,441  
; FILING DATE: 24-JUN-1994  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Feiber, Donna M.  
; REGISTRATION NUMBER: 33,878  
; REFERENCE/DOCKET NUMBER: 21-93C  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (303) 499-8080  
; TELEFAX: (303) 499-8089  
; INFORMATION FOR SEQ ID NO: 9:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 7266 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: double  
; TOPOLOGY: linear  
; MOLECULE TYPE: DNA (genomic)  
; HYPOTHETICAL: NO  
; ANTI-SENSE: NO  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: 949..6063  
; US-09-490-931-9  
  
Alignment Scores:  
Pred. No.: 1,72e+03 Length: 7266  
Score: 41.00 Matches: 7  
Percent Similarity: 72.73% Conservative: 1  
Best Local Similarity: 63.64% Mismatches: 3  
Query Match: 45.56% Indels: 0  
DB: 4 Gaps: 0  
  
US-10-020-441-3 (1-16) x US-09-490-931-9 (1-7266)  
Qy 3 TyrglyArgAspAspAsnAspArgAspProser 13  
Db 6380 TATGACGGGACTCGAACCCGCGACCCCTGCG 6348  
  
RESULT 38  
US-09-221-017B-473  
; Sequence 473, Application US/09221017B  
; Patent No. 6444799  
; GENERAL INFORMATION:  
; APPLICANT: Ross, Bruce C.  
; TITLE OF INVENTION: P. GINGIVALIS NUCLEOTIDES AND USES THEREOF  
; NUMBER OF SEQUENCES: 1120  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: MORRISON & ROEBSTER  
; STREET: 755 PAGE MILL ROAD  
; CITY: Palo Alto

STATE: CA  
COUNTRY: USA  
ZIP: 94304-1018  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: Windows  
SOFTWARE: FastSeq for Windows Version 2.0b  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/221,017B  
FILING DATE: 23-DEC-1998  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: PP1182  
FILING DATE: 31-DEC-1997  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: PP1546  
FILING DATE: 30-JAN-1998  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: PP2911  
FILING DATE: 09-APR-1998  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: PCT/AU98/01023  
FILING DATE: 10-DEC-1998  
ATTORNEY/AGENT INFORMATION:  
NAME: Monroy, Gladys H  
REGISTRATION NUMBER: 32,430  
REFERENCE/DOCKET NUMBER: 27340-20021.00  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 650-813-5600  
TELEFAX: 650-494-0792  
TELEX: 706141  
INFORMATION FOR SEQ ID NO: 473:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 8439 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: circular  
MOLECULE TYPE: DNA (genomic)  
HYPOTHEICAL: NO  
ANTI-SENSE: UNKNOWN  
ORIGINAL SOURCE:  
ORGANISM: PORYPHYROMONAS GINGIVALIS  
FEATURE:  
NAME/KEY: misc feature  
LOCATION: 1...8439  
US-09-221-017B-473

Alignment Scores:  
Pred. No.: 2.03e+03 Length: 8439  
Score: 41.00 Matches: 7  
Percent Similarity: 72.73% Conservative: 1  
Best Local Similarity: 63.64% Mismatches: 3  
Query Match: 45.56% Indels: 0  
DB: 4 Gaps: 0

US-10-020-441-3 (1-16) x US-09-221-017B-473 (1-8439)

Qy 3 TyrglyArGaPAspAmaSPArGaSPProSer 13  
Db 742 TATGACGCGACTCGAACCCGCGACCCCTGCG 774

RESULT 39  
US-08-570-311-28/c  
Sequence 28, Application US/08570311  
GENERAL INFORMATION:  
APPLICANT: Proguileke-Fox, Ann  
APPLICANT: Tumwasorn, Somying  
APPLICANT: Lepine, Guylaine  
APPLICANT: Han, Naiming  
APPLICANT: Lantz, Marilyn  
APPLICANT: Patti, Joseph

TITLE OF INVENTION: Cloned Porphyromonas gingivalis Genes  
TITLE OF INVENTION: and Probes for the Detection of Periodontal Disease  
NUMBER OF INVENTION: 29  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Ted W. Whitlock  
STREET: 2421 N.W. 41st Street, Suite A-1  
CITY: Gainesville  
STATE: FL  
COUNTRY: USA  
ZIP: 32606  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/570,311  
FILING DATE:  
CLASSIFICATION: 424  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/353,485  
FILING DATE: 09-DEC-1994  
CLASSIFICATION: 424  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/647,119  
FILING DATE: 25-JAN-1991  
CLASSIFICATION: 424  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/241,640  
FILING DATE: 08-SEP-1988  
ATTORNEY/AGENT INFORMATION:  
NAME: Whitlock, Ted W.  
REGISTRATION NUMBER: 36,965  
REFERENCE/DOCKET NUMBER: UP15.C3  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (904) 375-8100  
TELEFAX: (904) 372-5800  
INFORMATION FOR SEQ ID NO: 28:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 8640 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 971..6031  
US-08-570-311-28

Alignment Scores:  
Pred. No.: 2.09e+03 Length: 8640  
Score: 41.00 Matches: 7  
Percent Similarity: 72.73% Conservative: 1  
Best Local Similarity: 63.64% Mismatches: 3  
Query Match: 45.56% Indels: 0  
DB: 1 Gaps: 0

US-10-020-441-3 (1-16) x US-08-570-311-28 (1-8640)

Qy 3 TyrglyArGaPAspAmaSPArGaSPProSer 13  
Db 6350 TATGACGCGACTCGAACCCGCGACCCCTGCG 6318

RESULT 40  
US-09-734-673-3  
Sequence 3, Application US/09734673  
Patent No. 6410294  
GENERAL INFORMATION:  
APPLICANT: GUGLIER, Karl et al  
TITLE OF INVENTION: ISOLATED HUMAN KINASE PROTEINS, NUCLEIC  
ACID MOLECULES ENCODING HUMAN KINASE PROTEINS, AND USES  
TITLE OF INVENTION: THEREOF  
FILE REFERENCE: C1001020

```
; CURRENT APPLICATION NUMBER: US/09/734,673
; CURRENT FILING DATE: 2000-12-13
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 3
; LENGTH: 38564
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1)..(38564)
; OTHER INFORMATION: n = A,T,C or G
US-09-734-673-3
```

```
Alignment Scores:
Pred. No.: 1.12e+04 Length: 38564
Score: 41.00 Matches: 6
Percent Similarity: 76.92% Conservative: 4
Best Local Similarity: 46.15% Mismatches: 3
Query Match: 45.56% Indels: 0
DB: 4 Gaps: 0
```

US-10-020-441-3 (1-16) x US-09-734-673-3 (1-38564)

```
QY      4 G1YARGASPSPASnSPARGASPProSerARGLYSAsn 16
      |||||::|||:::|||||:::|||||
Db 31429 GGAAGGATGATGAGAGAGATGACCGAGGTAGAGAAAT 31467
```

Search completed: April 8, 2003, 19:53:42  
Job time : 22.1311 secs



GenCore version 5.1.4 p5 4578  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: April 8, 2003, 10:21:07 ; Search time 32 Seconds  
(without alignments)

1125.999 Million cell updates/sec

Title: US-10-020-441-3

Perfect score: 90

Sequence: 1 VGYGRDNDNDPDRKN 16

Scoring table:  
BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Xgapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 2185239 seqs, 112599159 residues

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Command line parameters:  
-MODEL=frame+ p2n model -DEV=xlh  
-Q=/cgm2\_1/USPFO.spool/US1002041/rumat\_02042003\_093011\_20959/app\_query.fasta\_1.462  
-DB=N\_Geneseq\_101002 -QFMT=fastap -SUFFIX=ring -MINMATCH=0.1 -LOOPEL=0  
-LOOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=blomemb2 -TRANS=human40.cdi  
-LIST=45 -DOCALLIGN=200 -THR SCORE=PCT -THR MAX=100 -THR MIN=0 -ALIGN=40  
-MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000  
-USER=US1002041 @CGN.1.1.275 @rumat\_02042003\_093011\_20959 -NCPU=6 -ICPU=3  
-NO\_XLPRY -NO\_MMAT -LARGEQUERY -NEG\_SCORES=0 -WAIT -LONGLOG -DEV TIMEOUT=120  
-WARN\_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -Fgapop=6 -Fgapext=7  
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :  
N\_Geneseq\_101002:\*  
1: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1980.DAT:\*  
2: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1981.DAT:\*  
3: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1982.DAT:\*  
4: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1983.DAT:\*  
5: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1984.DAT:\*  
6: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1985.DAT:\*  
7: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1986.DAT:\*  
8: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1987.DAT:\*  
9: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1988.DAT:\*  
10: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1989.DAT:\*  
11: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1990.DAT:\*  
12: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1991.DAT:\*  
13: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1992.DAT:\*  
14: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT:\*  
15: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1994.DAT:\*  
16: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1995.DAT:\*  
17: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT:\*  
18: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT:\*  
19: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT:\*  
20: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT:\*  
21: /SID52/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT:\*  
22: /SID52/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT:\*  
23: /SID52/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT:\*  
24: /SID52/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	90	100.0	1477	21	AAZ90608
2	52	57.8	2112	20	AAZ35744
3	52	57.8	2690	24	ABK35741
4	52	57.8	2833	20	AAZ06367
5	52	57.8	2833	22	AAZ0513
6	52	57.8	2880	20	AAZ68457
7	52	57.8	3492	20	AAZ35745
8	48	53.3	695	24	ABO29190
9	48	53.3	695	24	ABO29191
10	48	53.3	7906	22	AAZ45191
11	48	53.3	7906	24	ABN80179
12	48	53.3	7906	24	ABK28236
13	47	52.2	856	24	ABO19274
14	47	52.2	856	24	ABO19275
15	47	52.2	1742	21	AAZ37851
16	47	52.2	2275	22	AAZ31140
17	47	52.2	2275	24	ABO72512
18	47	52.2	2288	24	ABK71715
19	46	51.1	86	22	AAZ3527
20	46	51.1	138	22	AAZ70368
21	46	51.1	138	22	AAZ18605
22	46	51.1	138	22	AAZ44526
23	46	51.1	138	22	AAZ10515
24	46	51.1	138	24	ABZ18763
25	46	51.1	410	22	ABZ57751
26	46	51.1	410	22	AAZ05816
27	46	51.1	410	22	AAZ13443
28	46	51.1	410	22	AAZ13332
29	46	51.1	410	24	ABZ06197
30	46	51.1	1047	24	ABO26010
31	46	51.1	1047	24	ABO26011
32	46	51.1	8446	23	ABZ03976
33	46	50.0	500	21	AAZ19966
34	45	50.0	504	24	ABO14358
35	45	50.0	504	24	ABO14359
36	45	50.0	505	24	ABO32924
37	45	50.0	505	24	ABO32925
38	45	50.0	589	24	ABO53130
39	45	50.0	589	24	ABO53131
40	45	50.0	2154	21	AAZ48833
41	45	50.0	3675	23	AAZ80035
42	45	50.0	6200	24	ABZ42134
43	45	50.0	6216	24	ABZ10139
44	45	50.0	6216	24	ABZ39932
45	44	48.9	751	20	AAZ16323

ALIGNMENTS

RESULT 1  
AAZ90608  
ID AAZ90608 standard; DNA; 1477 BP.  
XX  
AC AAZ90608;  
XX  
DT 19-JUN-2000 (first entry)  
XX  
DE S. mansoni cercarial elastase gene.  
XX  
KW Schistosoma parasite; fusion protein; cercarial elastase; vaccine;  
XX antibody response; schistosomicide; ds.  
XX  
XX Schistosoma mansoni.  
XX  
Key Location/Qualifiers  
FH 1..1477  
FT CDS /\*tag= a

FT	/note="contains introns"
FT	1..168
FT	/*tag= b
FT	/number= 1
FT	/note="encodes the protein fragment shown in AAY67588"
FT	169..309
FT	/*tag= c
FT	/number= 1
FT	310..627
FT	/*tag= d
FT	/number= 2
FT	/note="encodes the protein fragment shown in AAY67587"
FT	628..1234
FT	/*tag= e
FT	/number= 2
FT	1235..1477
FT	/*tag= f
FT	/number= 3
XX	
PN	EP992582-A2.
XX	
PD	12-APR-2000.
XX	
PE	05-OCT-1999; 99EP-0307832.
XX	
PR	07-OCT-1998; 98GB-0021821.
XX	
PA	(UYWA-) UNIV WALES BANGOR.
XX	
PI	Doenhoff M, Sayers J;
DR	WPI; 2000-259136/23.
XX	
DR	P-PSDB; AAY67587, AAY67588.
XX	
PT	New vaccine for treatment of Schistosoma infections contains a
PT	recombinant fusion protein comprising cercarial elastase fused
PT	to bacterial, phage or viral protein -
XX	
PS	Claim 2; Fig 1; 26pp; English.
XX	
CC	The invention provides a vaccine comprising a recombinant fusion protein
CC	capable of eliciting immunity against Schistosoma parasites. The fusion
CC	protein comprises the 27 or 28 kDa cercarial elastase sequence of
CC	S. mansoni or an active fragment, homologue or variant, fused to a
CC	bacterial, phage or viral protein. The vaccine containing the fusion
CC	protein has been found to induce a significantly increased antibody
CC	response against schistosoma infections, compared to the use of S.
CC	mansoni cercarial elastase in its native form. The present sequence
CC	represents the S. mansoni cercarial elastase gene.
XX	
SQ	Sequence 1477 BP; 348 A; 261 C; 402 G; 465 T; 1 other;
XX	
Alignment Scores:	
Pred. No.:	1.91e-06 Length: 1477
Score:	90.00 Matches: 16
Percent Similarity:	100.00% Conservative: 0
Best Local Similarity:	100.00% Mismatches: 0
Query Match:	100.00% Indels: 0
DB:	21 Gaps: 0
US-10-020-441-3 (1-16) x AA290608 (1-1477)	
QY	1 VALGIYTYGILYARGASPASPASnAspARGSAPProSeARGLYSAsn 16
Db	562 GTTGTATTGGAAGGATGATACACCGCTGATCCGTACCTAAGAAAT 609
RESULT 2	
ID	AA35744 standard; cDNA; 2112 BP.
XX	
NC	AA35744;
XX	
DT	09-JUL-1999 (first entry)

XX	cDNA encoding a protein identified by the signal sequence trap method.
DE	
XX	Signal sequence trap method; SST method; immunisation; inhibition;
KW	infection; allergy; cancer; regulation; tissue formation; tissue repair;
KM	infection activity; inhibit activity; chemokine activity;
KW	cytokine activity; blood coagulation regulation; agonist; antagonist;
KM	metabolic disorder; hormonal disorder; immune disorder;
KM	severe combined immunodeficiency; SCID; AIDS; thrombosis; cancer;
XX	wound; ss.
OS	Homo sapiens.
XX	
PN	WO9918126-A1.
PD	15-APR-1999.
XX	
PF	06-OCT-1998; 98WO-JP04514.
XX	
PR	07-OCT-1997; 97JP-0274674.
XX	
PA	(ONOY ) ONO PHARM CO LTD.
XX	
PI	Fukushima D, Shibayama S, Tada H,
DR	WPT, 1999-277254/23.
XX	P-PSTDB; AAY02383.
XX	
PT	Polypeptides identified by the signal sequence trap method from a
XX	human cDNA library
PS	Claim 4; Page 242-243; 281pp; Japanese.
XX	
CC	AAX35694-X35747 represent cDNA sequences that encode novel polypeptides
CC	(AAX02358-84) which are identified from a human placental cDNA library
CC	by the signal sequence trap (SST) method. The polypeptides have a
CC	broad range of physiological activity, including immunisation against
CC	and inhibition of infections, allergies and cancer; regulation of tissue
CC	formation and repair; activin/inhibin activity; chemokine/cytokine
CC	activity; blood coagulation regulation; and receptor/ligand agonist
CC	or antagonist activity. The polypeptides can be used for prevention
CC	and treatment of disorders including infections by bacteria, yeasts and
CC	viruses (including HIV) and protozoa, metabolic and hormonal disorders;
CC	immune disorders (including severe combined immunodeficiency (SCID)
CC	and AIDS; thrombosis; cancer; and traumatic or surgical wounds.
XX	
SQ	Sequence 2112 BP; 603 A; 427 C; 469 G; 613 T; 0 other;
	Alignment Scores:
	Pred. No.: Length: 2112
	Score: 52.00 Matches: 11
	Percent Similarity: 85.71% Conservative: 1
	Best Local Similarity: 78.57% Mismatches: 1
	Query Match: 57.76% Indels: 1
	Ds: 20 Gaps: 0
US-10-020-441-3 (1-16) X AAX35744 (1-2112)	
Oy	1 ValGlyTyrGlyArgAspAspAsnAsp-ArgaspProSer 13
	:::::
Db	1375 ATGGAGATATGTACTGCACGCAATGATCAGACGCCAAGT 1336
RESULT 3	
ABK35741/C	
ID	ABK35741 standard; cDNA; 2690 BP.
XX	
AC	ABK35741;
XX	
DT	08-MAY-2002 (first entry)
XX	
DE	cDNA sequence #132 encoding novel human secreted protein.
XX	
KW	Human secreted protein; hyperproliferative disorder; autoimmune disorder

KM immune deficiency disorder; blood disorder; inflammatory disorder;  
KM infectious disorder; allergic condition; neurodegenerative disorder;  
KM liver fibrosis; coagulation disorder; gene therapy; antitubercial;  
KM tumour; cancer; hepatotropic; immunosuppressive; antirheumatic; gene; ss.  
XX  
OS Homo sapiens.  
XX  
PN WO200177289-A2.  
XX  
PD 18-OCT-2001.  
XX  
PF 29-MAR-2001; 2001WO-US10232.  
XX  
PR 06-APR-2000; 2000US-195605P.  
XX  
PA (GEMY ) GENETICS INST INC.  
XX  
PI Jacobs K, McCoy JM, Lavalie ER, Collins-Racie LA, Evans C,  
PI Merberg D, Treacy M, Agostino M, Bowman MR, Spaulding V, Wong GG,  
PI Clark HF, Fectel K, Howes SH, Resnick RJ, Gulukota K, Graham JR;  
XX  
DR WPI; 2002-179322/23.  
XX  
XX Six hundred and twenty three polynucleotides derived from a variety of  
PT human tissue sources which encode secreted proteins, useful for  
PT treating immune deficiencies and disorders such as autoimmune disorders  
PT  
XX  
PS Claim 1; Page 151-152; 393pp; English.  
XX  
XX The present invention relates to the isolation of novel cDNA sequences  
CC which encode human secreted proteins. The cDNA sequences have been  
CC derived from a variety of human tissues. The invention also provides  
CC a method for producing proteins from these polynucleotide sequences.  
CC The proteins are useful for identifying compounds that modulate their  
CC activity and production. The sequences of the invention are  
CC useful for treating diseases such as hyperproliferative disorders  
CC (e.g. cancer), immune deficiency disorders (e.g. severe combined  
CC immunodeficiency (SCID), autoimmune disorders (e.g. multiple  
CC sclerosis), blood disorders (e.g. thrombocytopenia), inflammatory  
CC disorders (e.g. arthritis), infectious disorders (e.g. hepatitis),  
CC allergic conditions (e.g. asthma), neurodegenerative disorders (e.g.  
CC Alzheimer's disease), liver fibrosis, coagulation disorders (e.g.  
CC haemophilia), and tumours. The polynucleotide sequences of the  
CC invention are also useful in gene therapy. ABK35610-ABK36232 represent  
CC the cDNA sequences of the invention that encode for novel human  
CC secreted proteins.  
XX  
SQ Sequence 2690 BP; 811 A; 464 C; 513 G; 902 T; 0 other;

Alignment Scores:  
Pred. No.: 20.8 Length: 2690  
Score: 52.00 Matches: 11  
Percent Similarity: 85.71% Conservative: 1  
Best Local Similarity: 78.57% Mismatches: 1  
Query Match: 57.78% Indels: 1  
DB: 24 Gaps: 0

US-10-020-441-3 (1-16) x ABK35741 (1-2690)

QY 1 ValGlyTyrGlyArgAspAspAsnAsp-ArgAspProSer 13  
Db 596 ATGGGATATGCTACTGACGACATCATCATGAGACCCCAAGT 557

RESULT 4  
AAZ06367/C  
ID AAZ06367 standard; DNA; 2833 BP.  
XX  
XX AAZ06367;  
XX  
XX 26-OCT-1999 (first entry)  
XX  
XX Human receptor protein (HURP) 2 nucleotide sequence.

XX  
XX receptor; cancer; autoimmune disorder; inflammation;  
KM antagonist; cell surface protein; cell signalling;  
KM antibody; ds.  
XX  
XX  
OS Homo sapiens.  
XX  
XX Key Location/Qualifiers  
XX CDS 172..2292  
XX FT /\*tag= a  
XX FT /product= Human receptor protein 2  
XX  
XX WO9944375-A2.  
XX  
XX 19-AUG-1999.  
XX  
XX 05-FEB-1999; 99WO-US02572.  
XX  
XX 12-FEB-1998; 98US-0022939.  
XX  
XX (INCY-) INCYTE PHARM INC.  
XX  
XX Au-Young J, Bandman O, Baughn M, Corley NC, Guegler KJ;  
PI Hillman JL, Lal P, Shah P, Tang YT, Yue H;  
XX  
XX WPI; 1999-494536/41.  
XX  
XX P-PSDB; AAY15223.  
XX  
XX New human receptor proteins, used e.g. to treat, prevent and  
PT diagnose gastrointestinal and developmental disorders - and related  
PT nucleic acids, vectors, transformed cells, antibodies, agonists and  
PT antagonists  
XX  
PS Claim 11; Page 87-88; 94pp; English.  
XX  
XX This is the consensus sequence for Human receptor protein 2 (HURP-2). It  
CC was derived from the following overlapping and/or extended nucleic acid  
CC sequences; Incyte clones 150629 (FIBRANT01), 639489 (BRSTN0703), 878146  
CC (LUNGAST01), 995140 (KIDNUT01), and shotgun sequences SAAB00536  
CC and SAIA03775.  
CC HURP-2 is expressed in cancerous, inflamed, fetal/proliferating, and  
CC reproductive tissue. HURP-2 therefore appears to have a role in some  
CC forms of cancer, autoimmune/inflammatory disorders, reproductive  
CC disorders, and developmental disorders.  
CC This gives rise to the possibility of using an antagonist or an antibody  
CC of HURP to treat or prevent cancer or autoimmune/inflammatory disorders.  
XX  
SQ Sequence 2833 BP; 832 A; 530 C; 617 G; 854 T; 0 other;

Alignment Scores:  
Pred. No.: 22 Length: 2833  
Score: 52.00 Matches: 11  
Percent Similarity: 85.71% Conservative: 1  
Best Local Similarity: 78.57% Mismatches: 1  
Query Match: 57.78% Indels: 1  
DB: 20 Gaps: 0

US-10-020-441-3 (1-16) x AAZ06367 (1-2833)

QY 1 ValGlyTyrGlyArgAspAspAsnAsp-ArgAspProSer 13  
Db 1552 ATGGGATATGCTACTGACGACATCATCATGAGACCCCAAGT 1513

RESULT 5  
AAF80513/C  
ID AAF80513 standard; cDNA; 2833 BP.  
XX  
XX AAF80513;  
XX  
XX 08-JUN-2001 (first entry)  
XX  
XX Receptor #1 partial coding sequence.

KW Probe; microarray; cancer; immunopathology; neuropathology; ss.  
 XX Mus musculus.  
 OS  
 XX US6183968-B1.  
 PN  
 XX 06-FEB-2001.  
 PD  
 XX 25-MAR-1999; 99US-0276531.  
 PF  
 XX 27-MAR-1998; 98US-0079677.  
 PR  
 XX (INCY-) INCYTE PHARM INC.  
 PA  
 XX Bandman O, Lal P, Hillman JL, Yue H, Reddy R, Guegler KJ;  
 PI Baughn MR;  
 DR WPI; 2001-201999/20.  
 XX  
 XX Composition having probes which comprise part of gene sequence encoding  
 PT proteins associated with cell proliferation useful as hybridizable  
 PT array elements in Microarrays to monitor expression of target  
 PT polynucleotide -  
 PS  
 PS Claim 1; Columns 57-60; 104pp; English.  
 XX  
 XX The present invention relates to a composition comprising several  
 CC polynucleotide probes. Probes can be derived from the present sequence.  
 CC The probes are immobilised and are preferably useful as hybridisable  
 CC array elements in a microarray for monitoring the expression of several  
 CC polynucleotides. The microarray can be used in the diagnosis of cancers  
 CC such as adenocarcinoma, leukaemia, lymphoma, melanoma, myeloma, sarcoma  
 CC and teratocarcinoma etc., immunopathology such as AIDS, Addison's  
 CC disease, adult respiratory distress syndrome, allergies, anaemia, asthma,  
 CC atherosclerosis and bronchitis etc., neuropathology such as Alzheimer's  
 CC disease, amnesia, amyotrophic lateral sclerosis, bipolar disorder,  
 CC catatonia and cerebral neoplasms etc. The microarray can also be used to  
 CC investigate an individual's predisposition to a disease such as cancer,  
 CC immunopathology or neuropathology. Also, the microarray can be used for  
 CC investigating cellular response to infection, drug treatment etc. The  
 CC microarray can be used for diagnostics, prognostics and treatment  
 CC regimens, drug discovery and development, toxicological and  
 CC carcinogenicity studies, forensics, pharmacogenomics etc. The array can  
 CC also be used for monitoring disease progression.  
 CC  
 XX  
 SQ Sequence 2633 BP; 832 A; 530 C; 617 G; 854 T; 0 other;  
 Alignment Scores:  
 Pred. No.: 22 Length: 2833  
 Score: 52.00 Matches: 11  
 Percent Similarity: 85.71% Conservative: 1  
 Best Local Similarity: 78.57% Mismatches: 1  
 Query Match: 57.78% Indels: 1  
 DB: 22 Gaps: 0  
 US-10-020-441-3 (1-16) x AAF80513 (1-2833)  
 QY 1 VALGTYTGTGTAAGAspAspAspAsp-ArgAspProser 13  
 DB 1552 ATGGGATATGTGTAAGACGACAAATGATCAGAGCCCAAGT 1513  
 RESULT 6  
 AAV68457/c  
 ID AAV68457 standard; cDNA; 2880 BP.  
 XX  
 AC AAV68457;  
 XX  
 DT 19-MAR-1999 (first entry)  
 XX  
 DE Human hFrizzled-6 protein coding sequence.  
 XX  
 XX Human; hFrizzled-6; G-protein coupled receptor; 7TM receptor; infection;  
 KW cancer; diabetes; eating disorder; Parkinson's disease; osteoporosis;

KW Huntington's disease; acute heart failure; urinary retention; therapy;  
 KW angina pectoris; myocardial infarction; benign prostatic hypertrophy;  
 KW psychotic disorder; neurological disorder; ss.  
 OS  
 XX Homo sapiens.  
 FH Key Location/Qualifiers  
 FT CDS 234..2354  
 FT /\*cag= a  
 XX  
 PN EP882793-A2.  
 XX  
 XX 09-DEC-1998.  
 PD  
 XX 28-MAY-1998; 98EP-0304208.  
 PF  
 XX 09-DEC-1997; 97US-0987289.  
 PR 02-JUN-1997; 97US-0048240.  
 XX  
 XX (SMIK ) SMITHKLINE BEECHAM CORP.  
 PA  
 XX Hu E, Zhu Y;  
 PI  
 DR WPI; 1999-011650/02.  
 DR P-PSDB; AAW82353.  
 XX  
 XX New nucleic acid encoding human frizzled-6 protein - used to treat,  
 PT prevent and diagnose e.g. infections, pain, cancer, diabetes,  
 PT obesity, anorexia, bulimia, asthma, Parkinson's disease, heart  
 PT failure, hypertension and hypotension  
 XX  
 XX Claim 2; Page 7-8; 24pp; English.  
 XX  
 XX This sequence encodes the human frizzled-6 protein of the invention. The  
 CC protein is a member of the G-protein coupled receptor (also known as 7TM  
 CC receptor) family, and has homology to the murine frizzled-6 protein.  
 CC Cells containing the DNA are used to express the protein, which is useful  
 CC for therapy, in drug screens and to raise antibodies (Ab) (including use  
 CC in vaccines). The protein or its agonists, or the DNA, are used to treat  
 CC or prevent diseases requiring increased activity or expression of  
 CC hFrizzled-6, while antagonists are used similarly where reduction in  
 CC activity or expression of the protein is required. Suitable diseases for  
 CC treatment are bacterial, fungal, protozoal or viral infections  
 CC (particularly by human immune deficiency virus), pain, cancer, diabetes,  
 CC obesity, anorexia, bulimia, asthma, Parkinson's or Huntington's diseases,  
 CC acute heart failure, hypo- or hyper-tension, urinary retention,  
 CC osteoporosis, angina pectoris, myocardial infarction, ulcers, allergy,  
 CC benign prostatic hypertrophy and psychotic or neurological disorders  
 CC (e.g. anxiety, schizophrenia and manic depression). These diseases, or  
 CC susceptibility to them, can be diagnosed by detecting a mutation in a  
 CC gene encoding hFrizzled-6 and/or by measuring levels of the protein or  
 CC associated RNA in a patient sample. Fragments of the DNA are useful as  
 CC hybridisation probes for isolating genomic and related sequences, as  
 CC diagnostic reagents (in standard amplification or hybridisation assays)  
 CC and for chromosome identification. Ab are used as immunoassay reagents,  
 CC for therapy, to identify hFrizzled-6-expressing clones and for affinity  
 CC purification.  
 XX  
 SQ Sequence 2880 BP; 835 A; 543 C; 629 G; 873 T; 0 other;  
 Alignment Scores:  
 Pred. No.: 22.4 Length: 2880  
 Score: 52.00 Matches: 11  
 Percent Similarity: 85.71% Conservative: 1  
 Best Local Similarity: 78.57% Mismatches: 1  
 Query Match: 57.78% Indels: 1  
 DB: 20 Gaps: 0  
 US-10-020-441-3 (1-16) x AAV68457 (1-2880)  
 QY 1 VALGTYTGTGTAAGAspAspAspAsp-ArgAspProser 13  
 DB 1614 ATGGGATATGTGTAAGACGACAAATGATCAGAGCCCAAGT 1575



ABQ29191  
 ID ABQ29191 standard; DNA; 695 BP.  
 AC ABQ29191;  
 XX  
 DT 12-JUL-2002 (first entry)  
 DE Oligonucleotide for detecting cytosine methylation SEQ ID NO 15782.  
 XX  
 DE Human; cytosine methylation; 5'-CpG-3'; uracil; cytosine; diagnosis;  
 KM drug; side effect; cancer; central nervous system; cardiovascular;  
 KM gastrointestinal; respiratory system; single nucleotide polymorphism;  
 KM SNP; cell differentiation; ds.  
 XX  
 OS Homo sapiens.  
 PN WO200218632-A2.  
 XX  
 PD 07-MAR-2002.  
 PF 01-SEP-2001; 2001WO-EP10074.  
 XX  
 PF 01-SEP-2000; 2000DE-1043826.  
 PR 05-SEP-2000; 2000DE-1044543.  
 XX  
 PA (EPIG-) EPIGENOMICS AG.  
 XX  
 PI Olek A, Piepenbrock C, Berlin K, Guetig D;  
 DR WPI; 2002-371829/40.  
 XX  
 PT Determining the degree of cytosine methylation in genomic DNA, useful  
 PT for diagnosis and prognosis, comprises selective hybridization of  
 PT amplicons from chemically treated DNA -  
 XX  
 PS Claim 12; 56bp + Sequence Listing; 56bp; German.  
 XX  
 CC This invention describes a novel method for determining the degree of  
 CC methylation of a particular cytosine in a motif 5'-CpG-3', present in a  
 CC genomic sample of DNA. The sample is treated chemically to convert  
 CC cytosine (C) but not methylated C, to uracil, then part of the genomic  
 CC DNA that contains the target C is amplified to form a labeled amplicon.  
 CC The amplicon is hybridised to two classes, each with at least one  
 CC member, of oligonucleotides and/or peptide-nucleic acid (PNA) oligomers  
 CC and the degree of hybridisation to both classes is determined from the  
 CC label on the amplicon. From the ratio of labels hybridised to the two  
 CC classes of oligomers, the degree of methylation is calculated. The method  
 CC is used: (i) for diagnosis and/or prognosis of side effects of  
 CC therapeutic drugs and of a wide range of diseases, e.g. cancer, disorders  
 CC of the central nervous, cardiovascular, gastrointestinal and respiratory  
 CC systems etc., particularly by detecting mutations or single nucleotide  
 CC polymorphisms (SNP's); and (ii) for differentiation of cell or tissue  
 CC types and for investigating cell differentiation. The method allows the  
 CC methylation status of many C residues to be determined simultaneously.  
 CC ABQ13410-ABQ54121 represent genomic DNA sequences used to illustrate the  
 CC method for determining the degree of cytosine methylation described in  
 CC the disclosure of the invention.  
 XX  
 SQ Sequence 695 BP; 264 A; 257 C; 87 G; 87 T; 0 other;

Alignment Scores:  
 Pred. No.: 23.2 Length: 695  
 Score: 48.00 Matches: 8  
 Percent Similarity: 83.33% Conservative: 2  
 Best Local Similarity: 66.67% Mismatches: 2  
 Query Match: 53.33% Indels: 0  
 DB: 24 Gaps: 0

US-10-020-441-3 (1-16) x ABQ29191 (1-695)

QY 5 ArgAspAspAsnAspArgAspProSerArgIyAsn 16  
 DB 380 CGCGAGGAACGACCGCCAAACCCGCGTAAAC 415

RESULT 10  
 AAS45391/C  
 ID AAS45391 standard; DNA; 7906 BP.  
 AC AAS45391;  
 XX  
 DT 18-DEC-2001 (first entry)  
 DE Chemically pretreated complementary DNA associated with cell cycle #48.  
 XX  
 DE Cell cycle; human; CpG dinucleotide; cytosine methylation; HIV; aging;  
 KM human immunodeficiency virus; neurodegenerative disorder; solid tumour;  
 KM graft-versus-host disease; glomerular disease; Lewy body disease; cancer;  
 KM arthritis; arteriosclerosis; anti-HIV; neuroprotective; antiarthritic;  
 KM immunosuppressive; antitumour; cytostatic; antiarteriosclerotic; ds;  
 KM PCR primer.  
 XX  
 OS Homo sapiens.  
 PN WO200168911-A2.  
 XX  
 PD 20-SEP-2001.  
 PF 15-MAR-2001; 2001WO-EP02945.  
 XX  
 PF 15-MAR-2000; 2000DE-1013847.  
 PR 06-APR-2000; 2000DE-1019058.  
 PR 07-APR-2000; 2000DE-1019173.  
 PR 30-JUN-2000; 2000DE-1032529.  
 PR 01-SEP-2000; 2000DE-1043826.  
 XX  
 PA (EPIG-) EPIGENOMICS AG.  
 XX  
 PI Olek A, Piepenbrock C, Berlin K;  
 DR WPI; 2001-602751/68.  
 XX  
 PT Designing primers and probes for analysing diseases associated with  
 PT cytosine methylation state e.g. arthritis, cancer, aging,  
 PT arteriosclerosis comprising fragments of chemically modified genes  
 PT associated with cell cycle -  
 XX  
 PS Claim 1; SEQ ID NO 96; 28bp; English.  
 XX  
 CC Sequences AAS45296-AAS45520 represent chemically pretreated genomic DNA  
 CC molecules associated with the cell cycle and specific PCR primers of the  
 CC invention. The sequences are useful for detecting the methylation state  
 CC of all CpG dinucleotides in a sequence and therefore for analysing  
 CC associated diseases. By analysing cytosine methylations in the pretreated  
 CC DNA, genetic and/or epigenetic parameters for the diagnosis and therapy  
 CC of existing diseases or the predisposition to specific diseases can be  
 CC ascertained. The parameters may be compared to another set of genetic  
 CC and/or epigenetic parameters, the differences serving as basis for  
 CC diagnosis and/or prognosis events which are disadvantageous to patients.  
 CC The sequences of the invention are useful for the diagnosis and therapy  
 CC of HIV infection, neurodegenerative disorders, graft-versus-host disease,  
 CC aging, glomerular disease, Lewy body disease, arthritis,  
 CC arteriosclerosis, solid tumours and cancers.  
 XX  
 SQ Sequence 7906 BP; 1853 A; 310 C; 2142 G; 3601 T; 0 other;

Alignment Scores:  
 Pred. No.: 358 Length: 7906  
 Score: 48.00 Matches: 10  
 Percent Similarity: 81.25% Conservative: 3  
 Best Local Similarity: 62.50% Mismatches: 1  
 Query Match: 53.33% Indels: 2  
 DB: 22 Gaps: 1

US-10-020-441-3 (1-16) x AAS45391 (1-7906)

QY 3 TyGtIyArgAspAspAsnAspArg-----AspProSerArgIyAsn 16

DB 2985 TACCTACGCGACGACCAACGACCAACCAACCAACGACGCGGAAC 2938

RESULT 11

ID ABN80179 standard; DNA; 7906 BP.

AC ABN80179;

DT 15-JUL-2002 (first entry)

DE Human chemically modified disease associated gene SEQ ID NO 196.

XX Human; development; homeobox gene; HOX; diabetes; cancer; apoptosis;

KM heart disease; epilepsy; histone deacetylation; muscular dystrophy;

KM dwarfism; single nucleotide polymorphism; SNP; cytosine methylation;

XX antidiabetic; cytosine; anticonvulsant; de.

OS Homo sapiens.

XX Synthetic.

PN WO200200927-A2.

PD 03-JAN-2002.

XX 02-JUL-2001; 2001WO-EP07536.

PF 30-JUN-2000; 2000DE-1032529.

PR 01-SEP-2000; 2000DE-1043826.

XX (EPIG-) EPIGENOMICS AG.

XX Olek A, Piepenbrock C, Berlin K;

PI WPI; 2002-130908/17.

XX Novel nucleic acid useful for diagnosis and therapy of diseases

PT associated with development genes such as diabetes, comprises a

PT sequence of a segment of chemically pretreated DNA of genes associated

PT with development

XX Claim 1; SEQ ID NO 196; 27pp; English.

XX The invention relates to a nucleic acid (I) comprising a sequence at

CC least 18 bases in length of a segment of chemically pretreated DNA (II)

CC of genes associated with development selected from 87 genes listed in

CC the specification such as ACCPN, ADPN, or APDI and comprising one of 350

CC sequences (ABN79984-ABN80333) or their complements. The invention is

CC useful for the diagnosis or therapy of diseases associated with

CC development genes, in particular disease related to homeobox containing

CC genes (HOX), like diabetes, cancer, apoptosis related diseases, syndromes

CC associated with congenital heart disease, epilepsy, diseases related to

CC histone deacetylation, Currarino syndrome, diseases related with the

CC development of the brain and limb girdle muscular dystrophy and dwarfism.

CC Oligomers specific to each of the genes are useful for detecting the

CC methylation state of all CG dinucleotides within the 350 sequences or

CC (II) and their complementary sequences, as primer oligonucleotides for

CC the amplification of the 350 sequences, (II) and/or their complements and

CC as oligomer probes for detecting the cytosine methylation state and/or

CC single nucleotide polymorphisms (SNPs).

CC Note: The sequence data for this patent did not form part of the printed

CC specification but is based on sequence information supplied to Derwent by

CC the European Patent Office.

XX

XX Sequence 7906 BP; 1853 A; 310 C; 2142 G; 3601 T; 0 other;

SO

Alignment Scores:

Pred. No.: 358

Score: 48.00

Percent Similarity: 81.25%

Best Local Similarity: 62.50%

Query Match: 53.33%

DB: 24

Length: 7906

Matches: 10

Conservative: 3

Mismatches: 1

Indels: 2

Gaps: 1

US-10-020-441-3 (1-16) x ABN80179 (1-7906)

Qy 3 TyrGlyArgAspAspAspArg-----AspProSerArgLysAsn 16

DB 2985 TACCTACGCGACGACCAACGACCAACCAACCAACGACGCGGAAC 2938

RESULT 12

ID ABK28236 standard; DNA; 7906 BP.

AC ABK28236;

DT 23-APR-2002 (first entry)

DE DNA transcription associated complementary genomic DNA #55.

XX DNA transcription associated gene; peptide nucleic acid; PNA-oligomer;

KM PNA; cytosine methylation state; SNP; retroviral infection; gene; ds;

KM single nucleotide polymorphism; adenosine deaminase deficiency; cancer;

KM viral infection; Sezary syndrome; haematological disorder; tuberculosis;

KM immunological disorder; Werner syndrome; developmental disorder;

KM psoriasis; Rieger's syndrome; neurological disorder; erythropoiesis;

KM neurodegenerative disorder; Waardenburg syndrome; Niemann-Pick disease;

KM myelodysplastic syndrome; myocardial infarction; hypertension; arthritis;

KM angiodysplasia; congenital heart disease; HDR syndrome; gene therapy;

KM polyglutamine disorder; solid tumour.

XX Unidentified.

OS WO200192565-A2.

XX 06-APR-2001; 2001WO-EP03973.

XX 06-DEC-2001.

XX 06-APR-2000; 2000DE-1019058.

XX 07-APR-2000; 2000DE-1019173.

XX 30-JUN-2000; 2000DE-1032529.

XX 01-SEP-2000; 2000DE-1043826.

XX (EPIG-) EPIGENOMICS AG.

XX Olek A, Piepenbrock C, Berlin K;

PI WPI; 2002-090046/12.

XX New nucleic acids or oligomers, useful for diagnosing or treating

PT diseases associated with DNA transcription, e.g. immunological

PT disorders, Werner syndrome, psoriasis, myocardial infarction, solid

PT tumours or cancer

XX Claim 1; SEQ ID NO 110; 33pp; English.

XX The invention relates to a nucleic acid, which comprises a segment of the

CC chemically pretreated DNA of genes associated with DNA transcription from

CC one of 346 sequences, and an oligomer, in particular an oligonucleotide

CC or peptide nucleic acid (PNA)-oligomer that hybridises to or is identical

CC to the chemically pretreated DNA of genes associated with DNA

CC transcription. The set of oligomer probes are useful for detecting the

CC cytosine methylation state and/or single nucleotide polymorphisms (SNPs)

CC in a chemically pretreated genomic DNA. The nucleic acids are useful for

CC diagnosing or treating diseases associated with DNA transcription

CC (particularly with the methylation status), e.g. adenosine deaminase

CC deficiency, viral infection, retroviral infection, Sezary syndrome,

CC haematological disorders, immunological disorders, Werner syndrome,

CC tuberculosis, developmental disorders, psoriasis, Rieger's syndrome,

CC neurological disorders, neurodegenerative disorders, Waardenburg

CC syndrome, Niemann-Pick disease, myelodysplastic syndrome, myocardial

CC infarction, hypertension, angiodysplasia, erythropoiesis, congenital heart

CC disease, HDR syndrome, arthritis, polyglutamine disorders, solid tumours

CC or cancer. Sequences ABK28127-ABK28472 represent DNA transcription

CC associated genomic DNA molecules of the invention.





PR 14-MAY-1999; 99US-0134370.

PR	18-MAY-1999	99US-0134768
PR	18-MAY-1999	99US-0134768
PR	20-MAY-1999	99US-0135124
PR	21-MAY-1999	99US-0135523
PR	24-MAY-1999	99US-0135659
PR	25-MAY-1999	99US-0136691
PR	27-MAY-1999	99US-0136782
PR	28-MAY-1999	99US-0136782
PR	01-JUN-1999	99US-0137522
PR	03-JUN-1999	99US-0137528
PR	04-JUN-1999	99US-0137502
PR	07-JUN-1999	99US-0137724
PR	08-JUN-1999	99US-0138094
PR	10-JUN-1999	99US-0138547
PR	10-JUN-1999	99US-0138840
PR	14-JUN-1999	99US-0139191
PR	16-JUN-1999	99US-0139452
PR	17-JUN-1999	99US-0139493
PR	17-JUN-1999	99US-0139492
PR	18-JUN-1999	99US-0139454
PR	18-JUN-1999	99US-0139456
PR	18-JUN-1999	99US-0139455
PR	18-JUN-1999	99US-0139462
PR	18-JUN-1999	99US-0139463
PR	18-JUN-1999	99US-0139750
PR	18-JUN-1999	99US-0139763
PR	21-JUN-1999	99US-0139817
PR	22-JUN-1999	99US-0139899
PR	22-JUN-1999	99US-0140053
PR	23-JUN-1999	99US-0140054
PR	24-JUN-1999	99US-0140585
PR	28-JUN-1999	99US-0140823
PR	30-JUN-1999	99US-0140991
PR	01-JUL-1999	99US-0141287
PR	01-JUL-1999	99US-0141254
PR	02-JUL-1999	99US-0142055
PR	06-JUL-1999	99US-0142390
PR	08-JUL-1999	99US-0142803
PR	09-JUL-1999	99US-0142920
PR	12-JUL-1999	99US-0142927
PR	13-JUL-1999	99US-0143542
PR	14-JUL-1999	99US-0143624
PR	15-JUL-1999	99US-0144005
PR	16-JUL-1999	99US-0144086
PR	16-JUL-1999	99US-0144315
PR	19-JUL-1999	99US-0144331
PR	19-JUL-1999	99US-0144332
PR	19-JUL-1999	99US-0144332
PR	19-JUL-1999	99US-0144334
PR	19-JUL-1999	99US-0144335
PR	20-JUL-1999	99US-0144352
PR	20-JUL-1999	99US-0144352
PR	20-JUL-1999	99US-0144352
PR	20-JUL-1999	99US-0145276
PR	26-JUL-1999	99US-0145276
PR	27-JUL-1999	99US-0145518
PR	27-JUL-1999	99US-0145518

PR 27-JUL-1999; 99US-0145919.  
PR 28-JUL-1999; 99US-0145951.  
PR 02-AUG-1999; 99US-0146386.  
PR 02-AUG-1999; 99US-0146388.  
PR 02-AUG-1999; 99US-0146389.  
PR 03-AUG-1999; 99US-0147038.  
PR 04-AUG-1999; 99US-0147202.  
PR 05-AUG-1999; 99US-0147302.  
PR 05-AUG-1999; 99US-0147260.  
PR 06-AUG-1999; 99US-0147302.  
PR 06-AUG-1999; 99US-0147416.  
PR 09-AUG-1999; 99US-0147493.  
PR 09-AUG-1999; 99US-0147935.  
PR 10-AUG-1999; 99US-0148171.  
PR 11-AUG-1999; 99US-0148319.  
PR 12-AUG-1999; 99US-0148341.  
PR 13-AUG-1999; 99US-0148565.  
PR 13-AUG-1999; 99US-0148684.  
PR 16-AUG-1999; 99US-0149368.  
PR 17-AUG-1999; 99US-0149175.  
PR 18-AUG-1999; 99US-0149426.  
PR 20-AUG-1999; 99US-0149722.  
PR 20-AUG-1999; 99US-0149723.  
PR 20-AUG-1999; 99US-0149929.  
PR 23-AUG-1999; 99US-0149902.  
PR 23-AUG-1999; 99US-0149930.  
PR 25-AUG-1999; 99US-0150566.  
PR 26-AUG-1999; 99US-0150884.  
PR 27-AUG-1999; 99US-0151065.  
PR 27-AUG-1999; 99US-0151066.  
PR 27-AUG-1999; 99US-0151080.  
PR 30-AUG-1999; 99US-0151303.  
PR 31-AUG-1999; 99US-0151438.  
PR 01-SEP-1999; 99US-0151930.  
PR 07-SEP-1999; 99US-0152363.  
PR 10-SEP-1999; 99US-0153070.  
PR 13-SEP-1999; 99US-0153758.  
PR 15-SEP-1999; 99US-0154018.  
PR 16-SEP-1999; 99US-0154039.  
PR 20-SEP-1999; 99US-0154779.  
PR 22-SEP-1999; 99US-0155139.  
PR 23-SEP-1999; 99US-0155486.  
PR 24-SEP-1999; 99US-0155659.  
PR 28-SEP-1999; 99US-0156458.  
PR 29-SEP-1999; 99US-0156596.  
PR 04-OCT-1999; 99US-0157117.  
PR 05-OCT-1999; 99US-0157753.  
PR 06-OCT-1999; 99US-0157865.  
PR 07-OCT-1999; 99US-0158029.  
PR 08-OCT-1999; 99US-0158232.  
PR 12-OCT-1999; 99US-0158369.  
PR 13-OCT-1999; 99US-0158293.  
PR 13-OCT-1999; 99US-0159294.  
PR 13-OCT-1999; 99US-0159295.  
PR 14-OCT-1999; 99US-0159329.  
PR 14-OCT-1999; 99US-0159330.  
PR 14-OCT-1999; 99US-0159331.  
PR 14-OCT-1999; 99US-0159637.  
PR 14-OCT-1999; 99US-0159638.  
PR 18-OCT-1999; 99US-0159584.  
PR 21-OCT-1999; 99US-0160741.  
PR 21-OCT-1999; 99US-0160767.  
PR 21-OCT-1999; 99US-0160768.  
PR 21-OCT-1999; 99US-0160770.  
PR 21-OCT-1999; 99US-0160814.  
PR 21-OCT-1999; 99US-0160815.  
PR 22-OCT-1999; 99US-0160980.  
PR 22-OCT-1999; 99US-0160981.  
PR 22-OCT-1999; 99US-0160989.  
PR 25-OCT-1999; 99US-0161404.  
PR 25-OCT-1999; 99US-0161405.  
PR 25-OCT-1999; 99US-0161406.

PR 26-OCT-1999; 99US-0161359.  
PR 26-OCT-1999; 99US-0161360.  
PR 26-OCT-1999; 99US-0161361.  
PR 28-OCT-1999; 99US-0161920.  
PR 28-OCT-1999; 99US-0161982.  
PR 28-OCT-1999; 99US-0161993.  
PR 29-OCT-1999; 99US-0162142.  
  
Alignment Scores:  
Pred. No.: 98.3 Length: 1742  
Score: 47.00 Matches: 9  
Percent Similarity: 73.33% Conservative: 2  
Best Local Similarity: 60.00% Mismatches: 4  
Query Match: 52.22% Indels: 0  
DB: Gaps: 21  
  
US-10-020-441-3 (1-16) x AAC37851 (1-1742)  
  
QY 1 ValGIYrGIYArGAsPAsPAsnAsPArGAsPProSeArGLys 15  
Db 959 ATGGGCTATGGCAGAGATGATCAATGCGCCGCGTCGACAGAAAA 1003  
  
RESULT 16  
ID AAS31140 standard; cDNA; 2275 BP.  
XX AAS31140;  
AC AAS31140;  
XX 04-DEC-2001 (first entry)  
DT  
XX Human diagnostic and therapeutic polynucleotide (DITHP) #155.  
DE Human; receptor; diagnostic; therapeutic; gene therapy; vaccine;  
KW cell proliferative disorder; Crohn's disease; lymphoma; leukaemia;  
KW acquired immune deficiency syndrome; AIDS; autoimmune disorder;  
KW respiratory disorder; ss.  
XX  
OS Homo sapiens.  
XX  
PN WO200162927-A2.  
XX  
PD 30-AUG-2001.  
XX  
PF 21-FEB-2001; 2001WO-US06059.  
XX  
PR 24-FEB-2000; 2000US-0184693.  
PR 24-FEB-2000; 2000US-0184697.  
PR 24-FEB-2000; 2000US-0184698.  
PR 24-FEB-2000; 2000US-0184768.  
PR 24-FEB-2000; 2000US-0184769.  
PR 24-FEB-2000; 2000US-0184770.  
PR 24-FEB-2000; 2000US-0184771.  
PR 24-FEB-2000; 2000US-0184772.  
PR 24-FEB-2000; 2000US-0184773.  
PR 24-FEB-2000; 2000US-0184774.  
PR 24-FEB-2000; 2000US-0184776.  
PR 24-FEB-2000; 2000US-0184777.  
PR 24-FEB-2000; 2000US-0184779.  
PR 24-FEB-2000; 2000US-0184797.  
PR 24-FEB-2000; 2000US-0184813.  
PR 24-FEB-2000; 2000US-0184813.  
PR 24-FEB-2000; 2000US-0184837.  
PR 24-FEB-2000; 2000US-0184841.  
PR 24-FEB-2000; 2000US-0185213.  
PR 24-FEB-2000; 2000US-0185216.  
PR 12-MAY-2000; 2000US-0203785.  
PR 15-MAY-2000; 2000US-0204228.  
PR 16-MAY-2000; 2000US-0204525.  
PR 16-MAY-2000; 2000US-0204908.  
PR 16-MAY-2000; 2000US-0204908.  
PR 17-MAY-2000; 2000US-0204815.  
PR 17-MAY-2000; 2000US-0204863.  
PR 17-MAY-2000; 2000US-0205221.  
PR 17-MAY-2000; 2000US-0205285.

PR 17-MAY-2000; 2000US-0205286.  
 PR 17-MAY-2000; 2000US-0205287.  
 PR 17-MAY-2000; 2000US-0205323.  
 PR 17-MAY-2000; 2000US-0205324.  
 XX  
 PA (INCY-) INCYTE GENOMICS INC.  
 PI Panzer SR, Spiro PA, Banville SC, Shah P, Chalup MS, Chang SC;  
 PI Chen A, D'Sa SA, Amesley S, Dahl CR, Dam TC, Daniels SE;  
 PI Dufour GE, Flores V, Fong WT, Greenwalt LB, Hillman JL, Jones AJ;  
 PI Liu TF, Roseberry AM, Rosen BH, Russo FD, Stockdreher TK, Datto A;  
 PI Wright RJ, Yap PE, Yu JY, Bradley DL, Bratcher SR, Chen W;  
 PI Cohen HJ, Hodgson DM, Lincoln SE, Jackson S;  
 XX  
 DR WPI: 2001-502867/55.  
 DR P-PSDB; AAU19569.  
 PT Polynucleotides encoding diagnostic and therapeutic proteins, e.g.  
 PT enzymes, hormones and receptors, useful in diagnostics and therapeutics  
 PT  
 PS  
 PS Claim 1; Page 369; 522pp; English.  
 XX  
 CC The invention relates to polynucleotides (I) encoding diagnostic and  
 CC therapeutic (DTRP) polypeptides (II), which include e.g. enzymes,  
 CC and proteins involved in growth and development and receptors. (I) and  
 CC (II) may be used in the prevention, diagnosis and treatment of diseases  
 CC associated with inappropriate DTRP expression. For example, (I) and  
 CC (II) may be used to treat disorders associated with decreased polypeptide  
 CC expression by rectifying mutations or deletions in a patient's genome,  
 CC that affect the activity of the DTRPs, by expressing inactive proteins  
 CC or supplementing the patient's own production of them. (I) and (II)  
 CC may be used to treat diseases, for example, cell proliferative disorder,  
 CC Crohn's disease, acquired immune deficiency syndrome (AIDS), lymphoma,  
 CC leukaemia, autoimmune disorders, and respiratory disorders. Additionally,  
 CC (I) may be used to produce the DTRPs, by inserting the nucleic acids  
 CC into a host cell and culturing the cell to express the protein. (I) and  
 CC its complementary sequences may also be used as DNA probes in diagnostic  
 CC assays to detect and quantitate the presence of similar nucleic acids in  
 CC samples, and therefore which patients may be in need of restorative  
 CC therapy. (II) may also be used as antigens in the production of  
 CC antibodies against DTRPs and in assays to identify modulators of DTRP  
 CC expression and activity. The anti-DTRP antibodies and antagonists may  
 CC also be used to down regulate expression and activity. The anti-DTRP  
 CC antibodies may also be used as diagnostic agents for detecting the  
 CC presence of DTRPs in samples (e.g. by enzyme linked immunosorbent  
 CC assay (ELISA)). AAS30986-AAS31196 represent human diagnostic and  
 CC therapeutic (DTRP) polynucleotides of the invention.  
 CC  
 XX  
 SQ Sequence 2275 BP; 697 A; 479 C; 597 G; 502 T; 0 other;  
 Alignment Scores:  
 Pred. No.: 133 Length: 2275  
 Score: 47.00 Matches: 8  
 Percent Similarity: 83.33% Conservative: 2  
 Best Local Similarity: 66.67% Mismatches: 2  
 Query Match: 52.22% Indels: 0  
 DB: 22 Gaps: 0  
 US-10-020-441-3 (1-16) x AAS31140 (1-2275)  
 OY 5 ArgAspAspAsnAspArgAspProSerArgLysAsn 16  
 DB 1704 CGAGGAGTGAGGACCGAGATCCCTCCATGAAAGAC 1739  
 RESULT 17  
 ID ABQ72512 standard; cDNA; 2275 BP.  
 XX  
 AC ABQ72512;  
 DT 03-SEP-2002 (first entry)  
 XX

DE Human MDR encoding cDNA SEQ ID NO 64.  
 XX  
 XX Human: MDR; disease detection and treatment molecule polynucleotide;  
 KW proliferative disorder; hepatitis; psoriasis; cancer; AIDS;  
 KW autoimmune disorder; inflammatory disorder; allergy; multiple sclerosis;  
 KW rheumatoid arthritis; transgenic; gene therapy; antiatherosclerotic;  
 KW hepatocytic; antiinflammatory; antipsoriatic; cyclostatic; anti-HIV;  
 KW antiallergic; antineoplastic; antiasthmatic; antiatherosclerotic; antitumor;  
 KW neuroprotective; antineuritic; antineuritic; gene; ss.  
 OS  
 XX Homo sapiens.  
 PN WO200240715-A2.  
 XX  
 PD 23-MAY-2002.  
 XX  
 XX  
 PF 06-SEP-2001; 2001WO-US27628.  
 XX  
 XX 06-SEP-2000; 2000US-230505P.  
 PR 06-SEP-2000; 2000US-230514P.  
 PR 06-SEP-2000; 2000US-230515P.  
 PR 06-SEP-2000; 2000US-230517P.  
 PR 06-SEP-2000; 2000US-230518P.  
 PR 06-SEP-2000; 2000US-230519P.  
 PR 06-SEP-2000; 2000US-230595P.  
 PR 06-SEP-2000; 2000US-230597P.  
 PR 06-SEP-2000; 2000US-230598P.  
 PR 06-SEP-2000; 2000US-230599P.  
 PR 06-SEP-2000; 2000US-230610P.  
 PR 06-SEP-2000; 2000US-230865P.  
 PR 06-SEP-2000; 2000US-230989P.  
 PR 06-SEP-2000; 2000US-230989P.  
 PR 07-SEP-2000; 2000US-230951P.  
 PR 07-SEP-2000; 2000US-231163P.  
 PR 07-SEP-2000; 2000US-231167P.  
 PR  
 PA (INCY-) INCYTE GENOMICS INC.  
 XX  
 XX Jackson S, Lincoln SE, Altus CM, Dufour GE, Chalup MS, Hillman JL;  
 PI Jones AL, Yu UT, Wright RJ, Gietzen D, Liu TF, Yap PE, Dahl CR;  
 PI Montlyama MG, Bradley DL, Rohatgi SD, Harris B, Roseberry AM;  
 PI Gerstein EH, Peralta CH, David MH, Panzer SR, Flores V, Datto A;  
 PI Marwaha R, Chen AJ, Chang SC, Au AP, Imman RR;  
 XX  
 DR WPI: 2002-527544/56.  
 DR P-PSDB; ABP51294.  
 XX  
 XX Novel human disease detection and treatment polypeptide, useful in  
 PT diagnosis, prevention or treatment of cell proliferative disorders e.g.  
 PT arteriosclerosis, cirrhosis and an autoimmune/inflammatory disorder  
 PT e.g. AIDS  
 XX  
 XX  
 XX Claim 1; Page 341-342; 618pp; English.  
 PS  
 PS The invention relates to an isolated human disease detection and  
 PS treatment (MDRP) polypeptide (I) selected from a polypeptide having a  
 CC sequence selected from 254 sequences (ABP51231-ABP51484) given in the  
 CC specification, a naturally occurring polypeptide comprising a sequence  
 CC having at least 90% identity to (I) or a biologically active or  
 CC immunogenic fragment of (I). (I) is useful for screening a compound for  
 CC effectiveness as an agonist or antagonist, for screening a compound that  
 CC specifically binds (I) or modulates the activity of (I) and for  
 CC preparing a polyclonal or monoclonal antibody by hybridoma technology.  
 CC Nucleic acids (II) (ABQ72449-ABQ72700) encoding (I) are useful for  
 CC screening a compound for effectiveness in altering expression of a target  
 CC polynucleotide comprising. Oligonucleotides and antibodies are useful for  
 CC detecting MDRP in a sample or for assessing toxicity of a test compound,  
 CC in a diagnostic test for a condition or a disease associated with the  
 CC expression of MDRP in a biological sample, for detecting (I) in a sample,  
 CC and for purifying (I) from a sample. A composition comprising (I), an  
 CC agonist or antagonist is useful for treating a disease or condition  
 CC associated with decreased or increased expression of functional MDRP.  
 CC (I) or (II) are useful for diagnosing, treating or preventing disorders

CC associated with aberrant expression of MDT1, where the disorders are  
 CC selected from a cell proliferative disorder such as arteriosclerosis,  
 CC cirrhosis, hepatitis, psoriasis, and cancer and an  
 CC autoimmune/inflammatory disorder such as AIDS, Addison's disease,  
 CC allergy, anaemia, asthma, atherosclerosis, gout, multiple sclerosis or  
 CC rheumatoid arthritis. (ii) are useful for creating knockin humanised  
 CC animals or transgenic animals to model human diseases. In somatic or  
 CC gemline gene therapy, to generate a transcript image of a tissue or cell  
 CC type, for detecting differences in the chromosomal location due to  
 CC translocation or inversion among normal, carrier or affected individuals  
 CC and as hybridisation probes for mapping naturally occurring genomic  
 CC sequences.

XX SQ Sequence 2275 BP; 697 A; 480 C; 597 G; 501 T; 0 other;

#### Alignment Scores:

Pred. No.:	133	Length:	2275
Score:	47.00	Matches:	8
Percent Similarity:	83.33%	Conservative:	2
Best Local Similarity:	66.67%	Mismatches:	2
Query Match:	52.22%	Indels:	0
DB:	24	Gaps:	0

US-10-020-441-3 (1-16) x AB072512 (1-2275)

Qy 5 ArgAspAspAsnAspArgAspProSerArgLysAsn 16

Db 1704 CGAGAGATGAGACCGAGATCCTCCATGAAGAAC 1739

#### RESULT 18

ID ABK71715 standard; cDNA; 2288 BP.

XX AC ABK71715;

XX DT 30-JUN-2002 (first entry)

XX DE Human dlthp polynucleotide #181.

XX KW Human; dlthp; diagnostic and therapeutic polynucleotide; gene; ss; bone;  
 KW cell proliferative disorder; cancer; tumour; autoimmune disorder; brain;  
 KW inflammatory disorder; viral infection; bacterial infection; seizure;  
 KW fungal infection; parasitic infections; developmental disorder; breast;  
 KW endocrine disorder; metabolic disorder; neurological disorder; cervix;  
 KW gastrointestinal disorder; transport disorder; gene therapy; kidney;  
 KW adrenal gland; bone marrow; lung; ovary; pancreas; prostate; spleen;  
 KW skin; testis; thymus.

XX OS Homo sapiens.

XX PN WO200220754-A2.

XX PD 14-MAR-2002.

PF 29-AUG-2001; 2001WO-US27127.

XX 05-SEP-2000; 2000US-229747P.

PR 05-SEP-2000; 2000US-229748P.

PR 05-SEP-2000; 2000US-229749P.

PR 05-SEP-2000; 2000US-229750P.

PR 05-SEP-2000; 2000US-229751P.

PR 05-SEP-2000; 2000US-230583P.

PR 06-SEP-2000; 2000US-230514P.

PR 06-SEP-2000; 2000US-230515P.

PR 06-SEP-2000; 2000US-230517P.

PR 06-SEP-2000; 2000US-230518P.

PR 06-SEP-2000; 2000US-230519P.

PR 06-SEP-2000; 2000US-230597P.

PR 06-SEP-2000; 2000US-230598P.

PR 06-SEP-2000; 2000US-230599P.

PR 06-SEP-2000; 2000US-230610P.

PR 06-SEP-2000; 2000US-230865P.  
 PR 06-SEP-2000; 2000US-230988P.  
 PR 07-SEP-2000; 2000US-230951P.  
 PR 07-SEP-2000; 2000US-231163P.  
 PR 07-SEP-2000; 2000US-231167P.

XX PA (INCYTE GENOMICS INC.

XX Stuart J, Lincoln SE, Altus CM, Dufour GE, Chalup MS, Hillman JL,  
 PI Jones AL, Yu JY, Wright RJ, Gierzen D, Liu TF, Yap PE, Dahl CR,  
 PI Montiyama MG, Bradley DL, Rohargi SD, Harris B, Roseberry AM;  
 PI Gerstin EH, Peralta CH, David MH, Panzer SR, Flores V, Dafio A;  
 PI Marwana R, Chen AJ, Chang SC, Au AP, Imman RR;

XX WPI; 2002-383054/41.

DR P-PSDB; ABG60124.

XX An isolated polynucleotide useful in diagnostics and therapeutics -

XX Claim 1; Page 493; 686pp; English.

CC The invention relates to human diagnostic and therapeutic (dlthp)  
 CC polynucleotides and their associated polypeptides (dlthp polypeptides).  
 CC The sequences of the invention are used in the treatment and diagnosis of  
 CC cell proliferative disorders (e.g. arteriosclerosis, cirrhosis), cancers  
 CC (e.g. tumours of the adrenal gland, bone, bone marrow, brain, breast,  
 CC cervix, kidney, lung, ovary, pancreas, prostate, skin, spleen, testis or  
 CC thymus), autoimmune/inflammatory disorders (e.g. asthma, bronchitis,  
 CC psoriasis, osteoporosis), viral infections, bacterial infections, fungal  
 CC infections, parasitic infections, developmental disorders (e.g. anaemia,  
 CC epilepsy), seizure disorders (e.g. cerebral palsy, spina bifida),  
 CC endocrine disorders (e.g. thrombosis, aneurysm), metabolic disorders  
 CC (e.g. obesity, diabetes), neurological disorders (e.g. stroke,  
 CC amyotrophic lateral sclerosis, multiple sclerosis), gastrointestinal  
 CC disorders (e.g. ulcerative colitis, lymphoma) and transport disorders  
 CC (e.g. myotonic dystrophy, catatonias, peripheral neuropathy). Sequences  
 CC ABK71535-ABK71809 represent human dlthp polynucleotides of the invention.

XX SQ Sequence 2288 BP; 710 A; 481 C; 596 G; 501 T; 0 other;

#### Alignment Scores:

Pred. No.:	134	Length:	2288
Score:	47.00	Matches:	8
Percent Similarity:	83.33%	Conservative:	2
Best Local Similarity:	66.67%	Mismatches:	2
Query Match:	52.22%	Indels:	0
DB:	24	Gaps:	0

US-10-020-441-3 (1-16) x ABK71715 (1-2288)

Qy 5 ArgAspAspAsnAspArgAspProSerArgLysAsn 16

Db 1704 CGAGAGATGAGACCGAGATCCTCCATGAAGAAC 1739

#### RESULT 19

ID AAS23527 standard; DNA; 86 BP.

XX AC AAS23527;

DT 04-DEC-2001 (first entry)

DE C. albicans essential gene knockout downstream PCR primer (KO-Down) #24.

XX KW Gene identification; essential gene; GRACE; pathogenic fungus;  
 KW gene replacement and conditional expression; fungal infection;  
 KW PCR primer; KO-Down; ss.

XX OS Candida albicans.

XX PN WO200160975-A2.

XX PD 23-AUG-2001.

XX 20-FEB-2001; 2001WO-US05551.  
 XX 18-FEB-2000; 2000US-0183534.  
 XX (ELIT-) ELITRA PHARM INC.  
 XX Roemer T, Jiang B, Boone C, Bussey H;  
 XX WPI; 2001-489080/53.  
 XX  
 XX Identifying genes essential to fungal metabolisms and identifying  
 XX potential therapeutic agents that target these genes -  
 XX  
 XX Disclosure; Page 268; 324pp; English.  
 XX  
 XX The present invention relates to novel methods for constructing fungal  
 XX strains useful for identification and validation of gene products as  
 XX targets for therapeutic agents, for creating a collection of identified  
 XX essential genes, and screening assays for the discovery of new drugs.  
 XX The invention provides the GRACE (gene replacement and conditional  
 XX expression) method for the construction of mutant organisms referred to  
 XX as GRACE strains of the organism. The invention can be applied to any  
 XX organism, particularly a pathogenic fungus e.g. Candida albicans,  
 XX Aspergillus fumigatus and Cryptococcus neoformans. The methods are  
 XX useful to identify agents that may be used in the treatment of fungal  
 XX infections. AAS23504-AAS23564 represent C. albicans essential gene  
 XX knockout downstream PCR primers (KO-Down) used in the methods of the  
 XX present invention.  
 XX  
 XX Sequence 86 BP; 34 A; 16 C; 17 G; 19 T; 0 other;

Alignment Scores:  
 Pred. No.: 5.02 Length: 86  
 Score: 46.00 Matches: 8  
 Percent Similarity: 90.00% Conservative: 1  
 Best Local Similarity: 80.00% Mismatches: 1  
 Query Match: 51.11% Indels: 0  
 DB: 22 Gaps: 0

US-10-020-441-3 (1-16) x AAS23527 (1-86)

Qy 5 ArgAspAspAspArgAspProSerArg 14  
 ID ABA70368/C  
 Db 45 AGAGCAACACGACGACCCCTCGAGG 74

RESULT 20  
 ABA70368 standard; DNA; 138 BP.  
 XX  
 XX ABA70368;  
 XX  
 XX 01-FEB-2002 (first entry)  
 XX  
 XX Human foetal liver single exon nucleic acid probe #18673.  
 XX  
 XX Human; foetal liver; gene expression; single exon nucleic acid probe; ss.  
 XX  
 XX Homo sapiens.  
 XX  
 XX WO200157277-A2.  
 XX  
 XX 09-AUG-2001.  
 XX  
 XX 30-JAN-2001; 2001WO-US00669.  
 XX  
 XX 04-FEB-2000; 2000US-0180312.  
 XX 26-MAY-2000; 2000US-0207456.  
 XX 30-JUN-2000; 2000US-0608408.  
 XX 03-AUG-2000; 2000US-0632366.  
 XX 21-SEP-2000; 2000US-0234687.  
 XX 27-SEP-2000; 2000US-0236359.  
 XX 04-OCT-2000; 2000GB-0024263.

XX (MOLE-) MOLECULAR DYNAMICS INC.  
 XX Penn SG, Hanzel DK, Chen W, Rank DR;  
 XX WPI; 2001-483447/52.  
 XX  
 XX Human genome-derived single exon nucleic acid probes useful for  
 XX analyzing gene expression in human fetal liver -  
 XX  
 XX Claim 4; SEQ ID NO 18673; 639bp + sequence listing; English.  
 XX  
 XX The invention relates to a single exon nucleic acid probe for  
 XX measuring human gene expression in a sample derived from human foetal  
 XX liver. The single exon nucleic acid probes may be used for predicting,  
 XX measuring and displaying gene expression in samples derived from human  
 XX foetal liver. The present sequence is a single exon nucleic acid  
 XX probe of the invention.  
 XX Note: The sequence data for this patent did not form part of the  
 XX printed specification, but was obtained in electronic format directly  
 XX from WIPO at ftp.wipo.int/pub/published\_pct\_sequences.  
 XX  
 XX Sequence 138 BP; 47 A; 35 C; 15 G; 41 T; 0 other;

Alignment Scores:  
 Pred. No.: 8.55 Length: 138  
 Score: 46.00 Matches: 9  
 Percent Similarity: 66.67% Conservative: 1  
 Best Local Similarity: 60.00% Mismatches: 5  
 Query Match: 51.11% Indels: 0  
 DB: 22 Gaps: 0

US-10-020-441-3 (1-16) x ABA70368 (1-138)

Qy 2 GlyTyrGlyArgAspAspAspArgAspProSerArgLysAsn 16  
 Db 123 GGTGGTGGTGGAGATGATGATGATGATGATCTAATAAAG 79

RESULT 21  
 AAK18605/C  
 ID AAK18605 standard; DNA; 138 BP.  
 XX  
 XX AAK18605;  
 XX  
 XX 05-NOV-2001 (first entry)  
 XX  
 XX Human brain expressed single exon probe SEQ ID NO: 18596.  
 XX  
 XX Human; brain expressed exon; gene expression analysis; probe;  
 XX microarray; Alzheimer's disease; multiple sclerosis; schizophrenia;  
 XX epilepsy; cancer; ss.  
 XX  
 XX Homo sapiens.  
 XX  
 XX WO200157275-A2.  
 XX  
 XX 09-AUG-2001.  
 XX  
 XX 30-JAN-2001; 2001WO-US00667.  
 XX  
 XX 04-FEB-2000; 2000US-0180312.  
 XX 26-MAY-2000; 2000US-0207456.  
 XX 30-JUN-2000; 2000US-0608408.  
 XX 03-AUG-2000; 2000US-0632366.  
 XX 21-SEP-2000; 2000US-0234687.  
 XX 27-SEP-2000; 2000US-0236359.  
 XX 04-OCT-2000; 2000GB-0024263.

XX (MOLE-) MOLECULAR DYNAMICS INC.  
 XX Penn SG, Hanzel DK, Chen W, Rank DR;  
 XX WPI; 2001-483446/52.

XX Single exon nucleic acid probes for analyzing gene expression in human  
PT brains -  
XX  
XX  
PS Example 4; SEQ ID NO: 18596; 650pp + Sequence Listing; English.  
XX  
CC The present invention provides a number of single exon nucleic acid  
CC probes which are derived from genomic sequences expressed in the human  
CC brain. They can be used to measure gene expression in brain cell samples,  
CC which may enable the diagnosis and improved treatment of nervous system  
CC diseases such as Alzheimer's disease, multiple sclerosis, schizophrenia,  
CC epilepsy and cancers. The present sequence is one of the probes of the  
CC invention.  
XX  
SQ Sequence 138 BP; 47 A; 35 C; 15 G; 41 T; 0 other;  
XX  
Alignment Scores:  
Pred. No.: 8.55 Length: 138  
Score: 46.00 Matches: 9  
Percent Similarity: 66.67% Conservative: 1  
Best Local Similarity: 60.00% Mismatches: 5  
Query Match: 51.11% Indels: 0  
DB: 22 Gaps: 0  
US-10-020-441-3 (1-16) x AAK18605 (1-138)  
OY 2 GLYTYRGLYARGASPASPAARGASPProSerArgLysAsn 16  
Db 123 GGTGGTGGTGGAGATGATGATGATGATGATGATTAATAAATAAC 79  
RESULT 22  
AAK44526/C  
ID AAK44526 standard; DNA; 138 BP.  
AC AAK44526;  
XX  
DT 06-NOV-2001 (first entry)  
XX  
DE Human bone marrow expressed single exon probe SEQ ID NO: 19083.  
XX  
KW Human; bone marrow expressed exon; gene expression analysis; probe;  
KW microarray; cancer; Leukaemia; Lymphoma; myeloma; ss.  
XX  
OS Homo sapiens.  
XX  
PN WO200157276-A2.  
XX  
PD 09-AUG-2001.  
XX  
PF 30-JAN-2001; 2001WO-US00668.  
XX  
PR 04-FEB-2000; 2000US-0180312.  
PR 26-MAY-2000; 2000US-0207456.  
PR 30-JUN-2000; 2000US-0608408.  
PR 03-AUG-2000; 2000US-0632366.  
PR 21-SEP-2000; 2000US-0234687.  
PR 27-SEP-2000; 2000US-0236359.  
PR 04-OCT-2000; 2000GB-0024263.  
XX  
PA (MOLE-) MOLECULAR DYNAMICS INC.  
XX  
PI Penn SG, Hanzel DK, Chen W, Rank DR;  
XX  
DR WPI; 2001-488900/53.  
XX  
PT Human genome-derived single exon nucleic acid probes useful for  
PT analyzing gene expression in human bone marrow -  
XX  
PS Example 4; SEQ ID NO: 19083; 658bp + Sequence Listing; English.  
XX  
CC The present invention provides a number of single exon nucleic acid  
CC probes which are derived from genomic sequences expressed in the human  
CC bone marrow. They can be used to measure gene expression in bone marrow

CC samples, which may enable the improved diagnosis and treatment of cancers  
CC such as lymphoma, leukemia and myeloma. The present sequence is one of  
CC the probes of the invention.  
XX  
SQ Sequence 138 BP; 47 A; 35 C; 15 G; 41 T; 0 other;  
XX  
Alignment Scores:  
Pred. No.: 8.55 Length: 138  
Score: 46.00 Matches: 9  
Percent Similarity: 66.67% Conservative: 1  
Best Local Similarity: 60.00% Mismatches: 5  
Query Match: 51.11% Indels: 0  
DB: 22 Gaps: 0  
US-10-020-441-3 (1-16) x AAK44526 (1-138)  
OY 2 GLYTYRGLYARGASPASPAARGASPProSerArgLysAsn 16  
Db 123 GGTGGTGGTGGAGATGATGATGATGATGATTAATAAATAAC 79  
RESULT 23  
AAI50515/C  
ID AAI50515 standard; DNA; 138 BP.  
XX  
AC AAI50515;  
XX  
DT 17-OCT-2001 (first entry)  
XX  
DE Probe #19201 used to measure gene expression in human placenta sample.  
XX  
KW Probe; microarray; human; placenta; antenatal diagnosis;  
KW genetic disorder; ss.  
XX  
OS Homo sapiens.  
XX  
PN WO200157272-A2.  
XX  
PD 09-AUG-2001.  
XX  
PF 30-JAN-2001; 2001WO-US00663.  
XX  
PR 04-FEB-2000; 2000US-0180312.  
PR 26-MAY-2000; 2000US-0207456.  
PR 30-JUN-2000; 2000US-0608408.  
PR 03-AUG-2000; 2000US-0632366.  
PR 21-SEP-2000; 2000US-0234687.  
PR 27-SEP-2000; 2000US-0236359.  
PR 04-OCT-2000; 2000GB-0024263.  
XX  
PA (MOLE-) MOLECULAR DYNAMICS INC.  
XX  
PI Penn SG, Hanzel DK, Chen W, Rank DR;  
XX  
DR WPI; 2001-488897/53.  
XX  
PT Human genome-derived single exon nucleic acid probes useful for  
PT analyzing gene expression in human placenta -  
XX  
PS Claim 25; SEQ ID NO 19201; 654bp; English.  
XX  
CC The present invention relates to single exon nucleic acid probes (SENPs).  
CC The present sequence is one such probe. The probes are useful for  
CC producing a microarray for predicting, measuring and displaying gene  
CC expression in samples derived from human placenta. The probes are useful  
CC for antenatal diagnosis of human genetic disorders.  
XX  
SQ Sequence 138 BP; 47 A; 35 C; 15 G; 41 T; 0 other;  
XX  
Alignment Scores:  
Pred. No.: 8.55 Length: 138  
Score: 46.00 Matches: 9  
Percent Similarity: 66.67% Conservative: 1  
Best Local Similarity: 60.00% Mismatches: 5

Query Match: 51.11% Indels: 0  
DB: 22 Gaps: 0  
US-10-020-441-3 (1-16) x AA150515 (1-138)  
Qy 2 G1YTYrG1YArGAsPAsPaSnaSPArGAsPProSeArG1YsAaSn 16  
Db 123 GGTGGTGGTGAAGATGATGATGATGATGATCTTAATATAAAC 79  
RESULT 24  
AB518763/C  
ID AB518763 standard; DNA, 138 BP.  
XX  
AC AB518763;  
XX  
DT 19-AUG-2002 (first entry)  
XX  
DE Human genome-derived single exon ORF from lung SEQ ID No 18754.  
XX  
Human; de; single exon probe; asthma; lung cancer; COPD; ILD;  
KM chronic obstructive pulmonary disease; interstitial lung disease;  
KM familial idiopathic pulmonary fibrosis; neurofibromatosis;  
KM tuberous sclerosis; Gaucher's disease; Niemann-Pick disease;  
KM Hermansky-Pudlak syndrome; sarcoidosis; pulmonary haemosiderosis;  
KM pulmonary histiocytosis; lymphangioleiomyomatosis; Karsenger syndrome;  
KM pulmonary alveolar proteinosis; fibrocystic pulmonary dysplasia;  
KM primary ciliary dyskinesia; pulmonary hypertension;  
KM hyaline membrane disease; open reading frame; ORF.  
XX  
OS Homo sapiens.  
XX  
PN WO200186003-A2.  
XX  
PD 15-NOV-2001.  
XX  
30-JAN-2001; 2001WO-US00665.  
XX  
PF 04-FEB-2000; 2000US-180312P.  
XX  
PR 26-MAY-2000; 2000US-207456P.  
XX  
PR 30-JUN-2000; 2000US-0608408.  
XX  
PR 03-AUG-2000; 2000US-0632366.  
XX  
PR 21-SEP-2000; 2000US-234687P.  
XX  
PR 27-SEP-2000; 2000US-236359P.  
XX  
PR 04-OCT-2000; 2000GB-0024263.  
XX  
PA (MOLE-) MOLECULAR DYNAMICS INC.  
XX  
PI Penn SG, Hanzel DK, Chen W, Rank DR;  
XX  
XX WPI; 2002-114183/15.  
XX  
PT Spatially-addressable set of single exon nucleic acid probes, used to  
XX  
PT measure gene expression in human lung samples -  
XX  
XX Claim 4; SEQ ID No 18754; 634pp; English.  
XX  
XX The invention relates to a spatially-addressable set of single exon  
XX  
XX nucleic acid probes for measuring gene expression in a sample derived  
XX  
XX from human lung comprising single exon nucleic acid probes having one of  
XX  
XX 12614 nucleic acid sequences mentioned in the specification, or their  
XX  
XX complements or the 12387 open reading frames derived from the 12614  
XX  
XX probes. Also included are a microarray comprising the novel set of  
XX  
XX probes; the novel set of probes which hybridize at high stringency to a  
XX  
XX nucleic acid expressed in the human lung; measuring gene expression in a  
XX  
XX sample derived from human lung, comprising (a) contacting the array with  
XX  
XX a collection of detectably labeled nucleic acids derived from human lung  
XX  
XX mRNA, and (b) measuring the label detectably bound to each probe of  
XX  
XX the array; identifying exons in a eukaryotic genome, comprising  
XX  
XX (a) algorithmically predicting at least one exon from genomic sequences  
XX  
XX of the eukaryote; and (b) detecting specific hybridisation of detectably  
XX  
XX labeled nucleic acids from eukaryote lung mRNA, to a single exon probe,  
XX  
XX having a fragment identical to the predicted exon, the probe is included  
XX  
XX in the above mentioned microarray; assigning exons to a single gene,

CC comprising (a) identifying exons from genomic sequence by the method  
CC above and (b) measuring the expression of each of the exons in several  
CC tissues and/or cell types using hybridisation to a single exon  
CC microarrays having a probe with the exon, where a common pattern of  
CC expression of the exons in the tissues and/or cell types indicates that  
CC the exons should be assigned to a single gene; a peptide comprising one  
CC of 12011 sequences, mentioned in the specification, or encoded by the  
CC probes/open reading frames (ORF). The probes are used for gene  
CC expression analysis, and for identifying exons in a gene, particularly  
CC using human lung derived mRNA and for the study of lung diseases  
CC such as asthma, lung cancer, chronic obstructive pulmonary disease  
CC (COPD), interstitial lung disease (ILD), familial idiopathic pulmonary  
CC fibrosis, neurofibromatosis, tuberous sclerosis, Gaucher's disease,  
CC Niemann-Pick disease, Hermansky-Pudlak syndrome, sarcoidosis, pulmonary  
CC haemosiderosis, pulmonary histiocytosis, lymphangioleiomyomatosis,  
CC pulmonary alveolar proteinosis, Karsenger syndrome, fibrocystic  
CC pulmonary dysplasia, primary ciliary dyskinesia, pulmonary hypertension  
CC and hyaline membrane disease. The present sequence is a single exon  
CC probe open reading frame of the invention.  
CC Note: The sequence data for this patent did not form part  
CC of the printed specification, but was obtained in electronic  
CC format directly from WIPO at  
CC ftp.wipo.int/pub/published\_pct\_sequences.  
XX  
SQ Sequence 138 BP; 47 A; 35 C; 15 G; 41 T; 0 other;  
XX  
Alignment Scores:  
Pred. No.: 8.55 Length: 138  
Score: 46.00 Matches: 9  
Percent Similarity: 66.67% Conservative: 1  
Best Local Similarity: 60.00% Mismatches: 5  
Query Match: 51.11% Indels: 0  
DB: 24 Gaps: 0  
US-10-020-441-3 (1-16) x AB518763 (1-138)  
Qy 2 G1YTYrG1YArGAsPAsPaSnaSPArGAsPProSeArG1YsAaSn 16  
Db 123 GGTGGTGGTGAAGATGATGATGATGATGATCTTAATATAAAC 79  
RESULT 25  
AB517751/C  
ID AB517751 standard; DNA, 410 BP.  
XX  
XX  
XX AB517751;  
XX  
AC  
XX  
DT 01-FEB-2002 (first entry)  
XX  
XX  
XX Human foetal liver single exon nucleic acid probe #6056.  
XX  
XX Human; foetal liver; gene expression; single exon nucleic acid probe; ss.  
XX  
XX Homo sapiens.  
XX  
XX WO200157277-A2.  
XX  
XX  
XX 09-AUG-2001.  
XX  
PD  
XX  
PF 30-JAN-2001; 2001WO-US00669.  
XX  
XX  
XX 04-FEB-2000; 2000US-0180312.  
XX  
XX 26-MAY-2000; 2000US-0207456.  
XX  
XX 30-JUN-2000; 2000US-0608408.  
XX  
XX 03-AUG-2000; 2000US-0632366.  
XX  
XX 21-SEP-2000; 2000US-0234687.  
XX  
XX 27-SEP-2000; 2000US-0236359.  
XX  
XX 04-OCT-2000; 2000GB-0024263.  
XX  
XX  
XX (MOLE-) MOLECULAR DYNAMICS INC.  
XX  
XX Penn SG, Hanzel DK, Chen W, Rank DR;  
XX  
XX WPI; 2001-483447/52.

```
XX Human genome-derived single exon nucleic acid probes useful for
PT analyzing gene expression in human fetal liver -
XX
PS Claim 1; SEQ ID NO 6056; 639pp + sequence listing; English.
XX
CC The invention relates to a single exon nucleic acid probe for
CC measuring human gene expression in a sample derived from human foetal
CC liver. The single exon nucleic acid probes may be used for predicting,
CC measuring and displaying gene expression in samples derived from human
CC fetal liver. The present sequence is a single exon nucleic acid
CC probe of the invention.
CC Note: The sequence data for this patent did not form part of the
CC printed specification, but was obtained in electronic format directly
CC from WIPO at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 410 BP; 152 A; 73 C; 61 G; 124 T; 0 other;

Alignment Scores:
Pred. No.: 29.1 Length: 410
Score: 46.00 Matches: 9
Percent Similarity: 66.67% Conservative: 1
Best Local Similarity: 60.00% Mismatches: 5
Query Match: 51.11% Indels: 0
DB: Gaps: 0

US-10-020-441-3 (1-16) x ABA57751 (1-410)

OY 2 G|T|T|G|T|A|G|A|S|P|A|S|P|A|S|P|A|S|P|P|S|E|R|A|G|L|Y|S|A|N 16
DB 400 GGTGTGTGTGAGATGATGATGATGATGATGATCTTATATAAAAC 356

RESULT 26
AAK05816/c
ID AAK05816 standard; DNA; 410 BP.
AC AAK05816;
XX
DT 05-NOV-2001 (first entry)
XX
DE Human brain expressed single exon probe SEQ ID NO: 5807.
XX
KW Human; brain expressed exon; gene expression analysis; probe;
KW microarray; Alzheimer's disease; multiple sclerosis; schizophrenia;
KW epilepsy; cancer; ss.
XX
OS Homo sapiens.
XX
PN WO200157275-A2.
XX
PD 09-AUG-2001.
XX
PF 30-JAN-2001; 2001WO-US00667.
XX
PR 04-FEB-2000; 2000US-0180312.
PR 26-MAY-2000; 2000US-0207456.
PR 30-JUN-2000; 2000US-0608408.
PR 03-AUG-2000; 2000US-0632366.
PR 21-SEP-2000; 2000US-0234687.
PR 27-SEP-2000; 2000US-0236359.
PR 04-OCT-2000; 2000GB-0024263.
XX
PA (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX
DR WPI; 2001-483446/52.
XX
PT Single exon nucleic acid probes for analyzing gene expression in human
PT brains -
XX
PS Example 4; SEQ ID NO: 5807; 650pp + Sequence Listing; English.
XX
```

```
CC The present invention provides a number of single exon nucleic acid
CC probes which are derived from genomic sequences expressed in the human
CC brain. They can be used to measure gene expression in brain cell samples,
CC which may enable the diagnosis and improved treatment of nervous system
CC diseases such as Alzheimer's disease, multiple sclerosis, schizophrenia,
CC epilepsy and cancers. The present sequence is one of the probes of the
CC invention.
XX
SQ Sequence 410 BP; 152 A; 73 C; 61 G; 124 T; 0 other;

Alignment Scores:
Pred. No.: 29.1 Length: 410
Score: 46.00 Matches: 9
Percent Similarity: 66.67% Conservative: 1
Best Local Similarity: 60.00% Mismatches: 5
Query Match: 51.11% Indels: 0
DB: Gaps: 0

US-10-020-441-3 (1-16) x AAK05816 (1-410)

OY 2 G|T|T|G|T|A|G|A|S|P|A|S|P|A|S|P|A|S|P|P|S|E|R|A|G|L|Y|S|A|N 16
DB 400 GGTGTGTGTGAGATGATGATGATGATGATGATCTTATATAAAAC 356

RESULT 27
AAK31443/c
ID AAK31443 standard; DNA; 410 BP.
AC AAK31443;
XX
DT 06-NOV-2001 (first entry)
XX
DE Human bone marrow expressed single exon probe SEQ ID NO: 6000.
XX
KW Human; bone marrow expressed exon; gene expression analysis; probe;
KW microarray; cancer; leukaemia; lymphoma; myeloma; ss.
XX
OS Homo sapiens.
XX
PN WO200157276-A2.
XX
PD 09-AUG-2001.
XX
PF 30-JAN-2001; 2001WO-US00668.
XX
PR 04-FEB-2000; 2000US-0180312.
PR 26-MAY-2000; 2000US-0207456.
PR 30-JUN-2000; 2000US-0608408.
PR 03-AUG-2000; 2000US-0632366.
PR 21-SEP-2000; 2000US-0234687.
PR 27-SEP-2000; 2000US-0236359.
PR 04-OCT-2000; 2000GB-0024263.
XX
PA (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX
DR WPI; 2001-488900/53.
XX
PT Human genome-derived single exon nucleic acid probes useful for
PT analyzing gene expression in human bone marrow -
XX
PS Example 4; SEQ ID NO: 6000; 658pp + Sequence Listing; English.
XX
CC The present invention provides a number of single exon nucleic acid
CC probes which are derived from genomic sequences expressed in the human
CC bone marrow. They can be used to measure gene expression in bone marrow
CC samples, which may enable the improved diagnosis and treatment of cancers
CC such as lymphoma, leukaemia and myeloma. The present sequence is one of
CC the probes of the invention.
XX
SQ Sequence 410 BP; 152 A; 73 C; 61 G; 124 T; 0 other;
XX
```







XX This invention describes a novel method for determining the degree of  
CC methylation of a particular cytosine in a motif 5'-CpG-3', present in a  
CC genomic sample of DNA. The sample is treated chemically to convert  
CC cytosine (C) but not methylated C, to uracil, then part of the genomic  
CC DNA that contains the target C is amplified to form a labeled amplicon.  
CC The amplicon is hybridised to two classes, each with at least one  
CC member, of oligonucleotides and/or peptide-nucleic acid (PNA) oligomers  
CC and the degree of hybridisation to both classes is determined from the  
CC label on the amplicon. From the ratio of labels hybridised to the two  
CC classes of oligomers, the degree of methylation is calculated. The method  
CC is used: (i) for diagnosis and/or prognosis of side effects of  
CC therapeutic drugs and of a wide range of diseases, e.g. cancer, disorders  
CC of the central nervous, cardiovascular, gastrointestinal and respiratory  
CC systems etc.; particularly by detecting mutations or single nucleotide  
CC polymorphisms (SNPs); and (ii) for differentiation of cell or tissue  
CC types and for investigating cell differentiation. The method allows the  
CC methylation status of many C residues to be determined simultaneously.  
CC ABL03410-AB034121 represent genomic DNA sequences used to illustrate the  
CC method for determining the degree of cytosine methylation described in  
CC the disclosure of the invention.

XX  
SQ Sequence 1047 BP, 428 A, 338 C, 128 G, 153 T, 0 other;

Alignment Scores:  
Pred. No.: 83.4 Length: 1047  
Score: 46.00 Matches: 8  
Percent Similarity: 90.91% Conservative: 2  
Best Local Similarity: 72.73% Mismatches: 1  
Query Match: 51.11% Indels: 0  
DB: 24 Gaps: 0

US-10-020-441-3 (1-16) x AB026011 (1-1047)

QY 5 ArgAspAaPaaAspAaPaaPaaProSeArGlyAsn 15  
Db 514 CGAAATCCGACGACCGATCCGACTCGAATA 546

RESULT 32  
ABL03976

ID ABL03976 standard; cDNA; 8446 BP.

XX ABL03976;

DT 26-MAR-2002 (first entry)

DE Drosophila melanogaster expressed polynucleotide SEQ ID NO 6410.

XX Drosophila; developmental biology; cell signalling; insecticide;

XX pharmaceutical; gene; ss.

OS Drosophila melanogaster.

PN WO200171042-A2.

PD 27-SEP-2001.

PF 23-MAR-2001; 2001WO-US09231.

PR 23-MAR-2000; 2000US-191637P.

PR 11-JUL-2000; 2000US-0614150.

PA (PEKE ) PE CORP NY.

PI Venter JC, Adams M, Li PWD, Myers EW;

XX WPI; 2001-656860/75.  
DR P-PSDB; ABB59873.  
XX  
XX New isolated nucleic acid detection reagent for detecting 1000 or more  
PT genes from Drosophila and for elucidating cell signalling and cell-cell  
PT interactions -  
XX

PS Claim 1; SEQ ID NO 6410; 21pp + Sequence Listing; English.

XX The invention relates to an isolated nucleic acid detection reagent  
CC capable of detecting 1000 or more genes from Drosophila. The invention is  
CC useful in developmental biology and in elucidating cell signalling and  
CC cell-cell interactions in higher eukaryotes for the development of  
CC insecticides, therapeutics and pharmaceutical drugs. The invention  
CC discloses genomic DNA sequences (AB116176-AB110511), expressed DNA  
CC sequences (AB101840-AB116175) and the encoded proteins  
CC (AAB57737-ABB72072).  
CC The sequence data for this patent did not form part of the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pct\_sequences.

XX  
SQ Sequence 8446 BP, 2459 A, 1579 C, 1831 G, 2577 T, 0 other;

Alignment Scores:  
Pred. No.: 872 Length: 8446  
Score: 46.00 Matches: 8  
Percent Similarity: 73.33% Conservative: 3  
Best Local Similarity: 53.33% Mismatches: 4  
Query Match: 51.11% Indels: 0  
DB: 23 Gaps: 0

US-10-020-441-3 (1-16) x ABL03976 (1-8446)

QY 2 GlyTyrGlyArgAspAaPaaAspAaPaaPaaProSeArGlyAsn 16  
Db 4668 GACATGATGATCAGATCGACTTGATATCTGATCTGATATGAGGAAT 4712

RESULT 33  
AAA71986

ID AAA71986 standard; DNA; 500 BP.

XX AAA71986;

DT 19-JAN-2001 (first entry)

DE Human VEGF-X PDGF-like domain DNA.

XX VEGF-X; vascular endothelial growth factor; human; vulnarary; cytostatic;  
KW antihumneumatic; antiarthritic; antidiabetic; treatment;  
KW angiogenesis regulator; vascularization regulator; cancer; psoriasis;  
KW rheumatoid arthritis; diabetic retinopathy; blood vessel; organ repair;  
KW tissue regeneration; tissue repair; wound; dermal ulcer; pressure sore;  
KW venous sore; diabetic ulcer; burns; skin graft growth; ds.

OS Homo sapiens.

XX Homo sapiens.

FT Key Location/Qualifiers  
FT CDS 15..464  
FT /\*tag= a  
FT /product= "VEGF-X PDGF-like domain"

PN WO200037641-A2.

PD 29-JUN-2000.

PE 21-DEC-1999; 99WO-US30503.

PR 22-DEC-1998; 98GB-0028377.

PR 18-MAR-1999; 99US-0124967.

PR 08-NOV-1999; 99US-0164131.

PA (JANC ) JANSSEN PHARM NV.

PI Gordon RD, Sprengel J, Yon JR, Dijkmans JH, Goslowska A;  
PI Dhanaraj SN, Xu J;

XX WPI; 2000-442669/38.  
DR P-PSDB; AAB10624.  
XX  
XX New vascular endothelial growth factor protein, useful for treating or

PT preventing diseases associated with inappropriate angiogenesis activity  
PT such as cancer, rheumatoid arthritis, psoriasis and wounds -  
XX  
PS Claim 43; Fig 24; 127pp; English.  
XX  
CC This invention describes a novel vascular endothelial growth factor-X  
CC (VEGF-X) protein (1a) and its encoding polynucleotide (1ia) which has  
CC (VEGF-X), cytostatic, antirheumatic, antiarthritic, antipsoriatic and  
CC (VEGF-X) activity and acts as an angiogenesis and vascularization  
CC regulator. An antisense molecule of the invention is useful for treating  
CC or preventing cancer, rheumatoid arthritis, psoriasis and diabetic  
CC retinopathy by inhibiting angiogenic activity or inappropriate  
CC vascularization including formation and proliferation of new blood  
CC vessels, growth and development of tissues, tissue regeneration and organ  
CC and tissue repair in a subject. The products of the invention are useful  
CC for preparing medicaments for treating wounds such as dermal ulcers,  
CC pressure sores, venous sores, diabetic ulcers and burns and to promote  
CC skin graft growth, tissue repair, proliferation of new blood vessels,  
CC tissue regeneration and organ repair by promoting angiogenic activity or  
CC vascularization. This sequence encodes a human VEGF-X protein PDGF-like  
CC domain which can be expressed in E. coli systems and which is described  
CC in the method of the invention.  
XX  
SQ Sequence 500 BP; 146 A; 106 C; 135 G; 113 T; 0 other;  
XX  
Alignment Scores:  
Pred. No.: 54.7 Length: 500  
Score: 45.00 Matches: 7  
Percent Similarity: 90.91% Conservative: 3  
Best Local Similarity: 63.64% Mismatches: 1  
Query Match: 50.00% Indels: 0  
DB: 21 Gaps: 0  
US-10-020-441-3 (1-16) x AAA71986 (1-500)  
QY 6 AspAspAaAspArgAspProSerArgLysAsn 16  
Db 93 GACGATGACGATACGATCCGCGAAGAAATCC 125  
RESULT 34  
ABQ14358/c  
ID ABQ14358 standard; DNA; 504 BP.  
XX  
AC ABQ14358;  
XX  
DT 12-JUL-2002 (first entry)  
XX  
DE Oligonucleotide for detecting cytosine methylation SEQ ID NO 949.  
XX  
KW Human; cytosine methylation; 5'-CpG-3'; uracil; cytosine; diagnosis;  
KW drug; side effect; cancer; central nervous system; cardiovascular;  
KW gastrointestinal; respiratory system; single nucleotide polymorphism;  
KW SNP; cell differentiation; ds.  
XX  
OS Homo sapiens.  
XX  
PN WO200218632-A2.  
XX  
PD 07-MAR-2002.  
XX  
PF 01-SEP-2001; 2001WO-EP10074.  
XX  
PR 01-SEP-2000; 2000DE-1043826.  
PR 05-SEP-2000; 2000DE-1044543.  
XX  
PA (EPiG-) EPIGENOMICS AG.  
XX  
PI Olek A, Piepenbrock C, Berlin K, Guetig D;  
XX  
DR WPI; 2002-371829/40.  
XX  
PT Determining the degree of cytosine methylation in genomic DNA, useful  
PT for diagnosis and prognosis, comprises selective hybridization of

PT amplicons from chemically treated DNA -  
XX  
XX Claim 12; 56pp + Sequence Listing; 56pp; German.  
XX  
CC This invention describes a novel method for determining the degree of  
CC methylation of a particular cytosine in a motif 5'-CpG-3', present in a  
CC genomic sample of DNA. The sample is treated chemically to convert  
CC cytosine (C) but not methylated C, to uracil, then part of the genomic  
CC DNA that contains the target C is amplified to form a labeled amplicon.  
CC The amplicon is hybridised to two classes, each with at least one  
CC member, of oligonucleotides and/or peptide-nucleic acid (PNA) oligomers  
CC and the degree of hybridization to both classes is determined from the  
CC label on the amplicon. From the ratio of labels hybridised to the two  
CC classes of oligomers, the degree of methylation is calculated. The method  
CC is used: (i) for diagnosis and/or prognosis of side effects of  
CC therapeutic drugs and of a wide range of diseases, e.g. cancer, disorders  
CC of the central nervous, cardiovascular, gastrointestinal and respiratory  
CC systems etc., particularly by detecting mutations or single nucleotide  
CC polymorphisms (SNP's) and (ii) for differentiation of cell or tissue  
CC types and for investigating cell differentiation. The method allows the  
CC methylation status of many C residues to be determined simultaneously.  
CC ABQ1410-ABQ54121 represent genomic DNA sequences used to illustrate the  
CC method for determining the degree of cytosine methylation described in  
CC the disclosure of the invention.  
XX  
SQ Sequence 504 BP; 75 A; 61 C; 188 G; 180 T; 0 other;  
XX  
Alignment Scores:  
Pred. No.: 55.2 Length: 504  
Score: 45.00 Matches: 8  
Percent Similarity: 83.33% Conservative: 2  
Best Local Similarity: 66.67% Mismatches: 2  
Query Match: 50.00% Indels: 0  
DB: 24 Gaps: 0  
US-10-020-441-3 (1-16) x ABQ14358 (1-504)  
QY 5 ArgAspAaAspArgAspProSerArgLysAsn 16  
Db 279 CCGGACGACAAACGACCGGACGACGCGAAGAACGAAAC 244  
RESULT 35  
ABQ14359  
ID ABQ14359 standard; DNA; 504 BP.  
XX  
AC ABQ14359;  
XX  
DT 12-JUL-2002 (first entry)  
XX  
DE Oligonucleotide for detecting cytosine methylation SEQ ID NO 950.  
XX  
KW Human; cytosine methylation; 5'-CpG-3'; uracil; cytosine; diagnosis;  
KW drug; side effect; cancer; central nervous system; cardiovascular;  
KW gastrointestinal; respiratory system; single nucleotide polymorphism;  
KW SNP; cell differentiation; ds.  
XX  
OS Homo sapiens.  
XX  
PN WO200218632-A2.  
XX  
PD 07-MAR-2002.  
XX  
PF 01-SEP-2001; 2001WO-EP10074.  
XX  
PR 01-SEP-2000; 2000DE-1043826.  
PR 05-SEP-2000; 2000DE-1044543.  
XX  
PA (EPiG-) EPIGENOMICS AG.  
XX  
PI Olek A, Piepenbrock C, Berlin K, Guetig D;  
XX  
DR WPI; 2002-371829/40.  
XX  
PT Determining the degree of cytosine methylation in genomic DNA, useful  
PT for diagnosis and prognosis, comprises selective hybridization of

PT Determining the degree of cytosine methylation in genomic DNA, useful  
 PT for diagnosis and prognosis, comprises selective hybridization of  
 PT amplicons from chemically treated DNA -  
 XX  
 PS Claim 12; 56pp + Sequence Listing; 56pp; German.  
 CC This invention describes a novel method for determining the degree of  
 CC methylation of a particular cytosine in a motif 5'-CpG-3', present in a  
 CC genomic sample of DNA. The sample is treated chemically to convert  
 CC cytosine (C) but not methylated C, to uracil, then part of the genomic  
 CC DNA that contains the target C is amplified to form a labeled amplicon.  
 CC The amplicon is hybridised to two classes, each with at least one  
 CC member, of oligonucleotides and/or peptide-nucleic acid (PNA) oligomers  
 CC and the degree of hybridisation to both classes is determined from the  
 CC label on the amplicon. From the ratio of labels hybridised to the two  
 CC classes of oligomers, the degree of methylation is calculated. The method  
 CC is used: (i) for diagnosis and/or prognosis of side effects of  
 CC therapeutic drugs and of a wide range of diseases, e.g. cancer, disorders  
 CC of the central nervous, cardiovascular, gastrointestinal and respiratory  
 CC systems etc., particularly by detecting mutations or single nucleotide  
 CC polymorphisms (SNP's); and (ii) for differentiation of cell or tissue  
 CC types and for investigating cell differentiation. The method allows the  
 CC methylation status of many C residues to be determined simultaneously.  
 CC ABQ13410-ABQ54121 represent genomic DNA sequences used to illustrate the  
 CC method for determining the degree of cytosine methylation described in  
 CC the disclosure of the invention.  
 XX  
 SQ Sequence 504 BP; 180 A; 188 C; 61 G; 75 T; 0 other;  
 Alignment Scores:  
 Pred. No.: 55.2 Length: 504  
 Score: 45.00 Matches: 8  
 Percent Similarity: 83.33% Conservative: 2  
 Best Local Similarity: 66.67% Mismatches: 2  
 Query Match: 50.00% Indels: 0  
 DB: 24 Gaps: 0  
 US-10-020-441-3 (1-16) x ABQ14359 (1-504)  
 Qy 5 ArgAaPaAaPaAaPaAaPaAaPProSeArGySaAn 16  
 Db 226 CGCGACGACAAAGACCGCGACTACGCGAAACGAAC 261  
 RESULT 36  
 ID ABQ32924 standard; DNA; 505 BP.  
 AC ABQ32924;  
 XX  
 DT 12-JUL-2002 (first entry)  
 XX  
 DE Oligonucleotide for detecting cytosine methylation SEQ ID NO 19515.  
 XX  
 KW Human; cytosine methylation; 5'-CpG-3'; uracil; cytosine; diagnosis;  
 KW drug; side effect; cancer; central nervous system; cardiovascular;  
 KW gastrointestinal; respiratory system; single nucleotide polymorphism;  
 KW SNP; cell differentiation; ds.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO200218632-A2.  
 XX  
 PD 07-MAR-2002.  
 XX  
 PF 01-SEP-2001; 2001WO-EP10074.  
 XX  
 PR 01-SEP-2000; 2000DE-1043826.  
 PR 05-SEP-2000; 2000DE-1044543.  
 XX  
 PA (EPIG-) EPIGENOMICS AG.  
 XX  
 PI Olek A, Piepenbrock C, Berlin K, Guetig D,  
 XX

DR WPI; 2002-371829/40.  
 XX  
 XX Determining the degree of cytosine methylation in genomic DNA, useful  
 PT for diagnosis and prognosis, comprises selective hybridization of  
 PT amplicons from chemically treated DNA -  
 XX  
 PS Claim 12; 56pp + Sequence Listing; 56pp; German.  
 CC This invention describes a novel method for determining the degree of  
 CC methylation of a particular cytosine in a motif 5'-CpG-3', present in a  
 CC genomic sample of DNA. The sample is treated chemically to convert  
 CC cytosine (C) but not methylated C, to uracil, then part of the genomic  
 CC DNA that contains the target C is amplified to form a labeled amplicon.  
 CC The amplicon is hybridised to two classes, each with at least one  
 CC member, of oligonucleotides and/or peptide-nucleic acid (PNA) oligomers  
 CC and the degree of hybridisation to both classes is determined from the  
 CC label on the amplicon. From the ratio of labels hybridised to the two  
 CC classes of oligomers, the degree of methylation is calculated. The method  
 CC is used: (i) for diagnosis and/or prognosis of side effects of  
 CC therapeutic drugs and of a wide range of diseases, e.g. cancer, disorders  
 CC of the central nervous, cardiovascular, gastrointestinal and respiratory  
 CC systems etc., particularly by detecting mutations or single nucleotide  
 CC polymorphisms (SNP's); and (ii) for differentiation of cell or tissue  
 CC types and for investigating cell differentiation. The method allows the  
 CC methylation status of many C residues to be determined simultaneously.  
 CC ABQ13410-ABQ54121 represent genomic DNA sequences used to illustrate the  
 CC method for determining the degree of cytosine methylation described in  
 CC the disclosure of the invention.  
 XX  
 SQ Sequence 505 BP; 77 A; 61 C; 198 G; 169 T; 0 other;  
 Alignment Scores:  
 Pred. No.: 55.3 Length: 505  
 Score: 45.00 Matches: 8  
 Percent Similarity: 83.33% Conservative: 2  
 Best Local Similarity: 66.67% Mismatches: 2  
 Query Match: 50.00% Indels: 0  
 DB: 24 Gaps: 0  
 US-10-020-441-3 (1-16) x ABQ32924 (1-505)  
 Qy 5 ArgAaPaAaPaAaPaAaPaAaPProSeArGySaAn 16  
 Db 370 CGCGACGACAAAGACCGCGACTACGCGAAACGAAC 335  
 RESULT 37  
 ID ABQ32925 standard; DNA; 505 BP.  
 AC ABQ32925;  
 XX  
 DT 12-JUL-2002 (first entry)  
 XX  
 DE Oligonucleotide for detecting cytosine methylation SEQ ID NO 19516.  
 XX  
 KW Human; cytosine methylation; 5'-CpG-3'; uracil; cytosine; diagnosis;  
 KW drug; side effect; cancer; central nervous system; cardiovascular;  
 KW gastrointestinal; respiratory system; single nucleotide polymorphism;  
 KW SNP; cell differentiation; ds.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO200218632-A2.  
 XX  
 PD 07-MAR-2002.  
 XX  
 PF 01-SEP-2001; 2001WO-EP10074.  
 XX  
 PR 01-SEP-2000; 2000DE-1043826.  
 PR 05-SEP-2000; 2000DE-1044543.  
 XX  
 PA (EPIG-) EPIGENOMICS AG.  
 XX



PR 05-SEP-2000; 2000DE-1044543.  
XX (EPIC-) EPIDEMIOLOGY AG.  
XX Olek A, Piepenbrock C, Berlin K, Guetig D;  
XX WPI; 2002-371829/40.  
XX  
XX Determining the degree of cytosine methylation in genomic DNA, useful  
PT for diagnosis and prognosis, comprises selective hybridization of  
PT amplicons from chemically treated DNA  
XX  
XX Claim 12; 56pp + Sequence Listing; 56pp; German.  
XX  
XX This invention describes a novel method for determining the degree of  
CC methylation of a particular cytosine in a motif 5'-CpG-3', present in a  
CC genomic sample of DNA. The sample is treated chemically to convert  
CC cytosine (C) but not methylated C, to uracil, then part of the genomic  
CC DNA that contains the target C is amplified to form a labeled amplicon.  
CC The amplicon is hybridized to two classes, each with at least one  
CC member, of oligonucleotides and/or peptide-nucleic acid (PNA) oligomers  
CC and the degree of hybridization to both classes is determined from the  
CC label on the amplicon. From the ratio of labels hybridized to the two  
CC classes of oligomers, the degree of methylation is calculated. The method  
CC is used: (i) for diagnosis and/or prognosis of side effects of  
CC therapeutic drugs and of a wide range of diseases, e.g. cancer, disorders  
CC of the central nervous, cardiovascular, gastrointestinal and respiratory  
CC systems etc., particularly by detecting mutations or single nucleotide  
CC polymorphisms (SNP's); and (ii) for differentiation of cell or tissue  
CC types and for investigating cell differentiation. The method allows the  
CC methylation status of many C residues to be determined simultaneously.  
CC AB013410-AB054121 represent genomic DNA sequences used to illustrate the  
CC method for determining the degree of cytosine methylation described in  
CC the disclosure of the invention.  
XX  
XX

SQ Sequence 589 BP; 201 A; 226 C; 84 G; 78 T; 0 other;

Alignment Scores:

Pred. No.:	65.8	Length:	589
Score:	45.00	Matches:	7
Percent Similarity:	90.00%	Conservative:	2
Best Local Similarity:	70.00%	Mismatches:	1
Query Match:	50.00%	Indels:	0
DB:	24	Gaps:	0

US-10-020-441-3 (1-16) x AB053131 (1-589)

QY 7 AsphanaaPaTgaPProSeArGlySaAn 16

Db 558 GAAACGACCGACCGAAGCGCGAAGC 587

RESULT 40

AAC48833 ID AAC48833 standard; DNA; 2154 BP.

XX AAC48833;

XX 18-OCT-2000 (first entry)

XX Arabidopsis thaliana DNA fragment SEQ ID NO: 58940.

XX Hybridisation assay; genetic mapping; gene expression control;

KM protein identification; signal transduction pathway;

KM metabolic pathway; promoter; termination sequence; ss.

XX Arabidopsis thaliana.

XX EP1033405-A2.

XX 06-SEP-2000.

XX 25-FEB-2000; 2000EP-0301439.

PR 25-FEB-1999; 99US-0121825.  
PR 05-MAR-1999; 99US-0123180.  
PR 09-MAR-1999; 99US-0123548.  
PR 23-MAR-1999; 99US-0125768.  
PR 25-MAR-1999; 99US-0126264.  
PR 29-MAR-1999; 99US-0126785.  
PR 01-APR-1999; 99US-0127462.  
PR 06-APR-1999; 99US-0128234.  
PR 08-APR-1999; 99US-0128714.  
PR 16-APR-1999; 99US-0129845.  
PR 19-APR-1999; 99US-0130077.  
PR 21-APR-1999; 99US-0130449.  
PR 23-APR-1999; 99US-0130510.  
PR 28-APR-1999; 99US-0130891.  
PR 30-APR-1999; 99US-0131449.  
PR 04-MAY-1999; 99US-0132048.  
PR 05-MAY-1999; 99US-0132407.  
PR 06-MAY-1999; 99US-0132486.  
PR 07-MAY-1999; 99US-0132487.  
PR 11-MAY-1999; 99US-0132863.  
PR 14-MAY-1999; 99US-0134256.  
PR 14-MAY-1999; 99US-0134219.  
PR 14-MAY-1999; 99US-0134221.  
PR 14-MAY-1999; 99US-0134370.  
PR 18-MAY-1999; 99US-0134768.  
PR 19-MAY-1999; 99US-0134941.  
PR 20-MAY-1999; 99US-0135124.  
PR 21-MAY-1999; 99US-0135353.  
PR 24-MAY-1999; 99US-0135629.  
PR 25-MAY-1999; 99US-0136021.  
PR 27-MAY-1999; 99US-0136382.  
PR 28-MAY-1999; 99US-0136782.  
PR 01-JUN-1999; 99US-0137222.  
PR 03-JUN-1999; 99US-0137528.  
PR 04-JUN-1999; 99US-0137502.  
PR 07-JUN-1999; 99US-0137724.  
PR 08-JUN-1999; 99US-0138094.  
PR 10-JUN-1999; 99US-0138540.  
PR 10-JUN-1999; 99US-0138847.  
PR 14-JUN-1999; 99US-0139119.  
PR 16-JUN-1999; 99US-0139452.  
PR 17-JUN-1999; 99US-0139453.  
PR 18-JUN-1999; 99US-0139454.  
PR 18-JUN-1999; 99US-0139455.  
PR 18-JUN-1999; 99US-0139456.  
PR 18-JUN-1999; 99US-0139457.  
PR 18-JUN-1999; 99US-0139458.  
PR 18-JUN-1999; 99US-0139459.  
PR 18-JUN-1999; 99US-0139460.  
PR 18-JUN-1999; 99US-0139461.  
PR 18-JUN-1999; 99US-0139462.  
PR 18-JUN-1999; 99US-0139463.  
PR 18-JUN-1999; 99US-0139750.  
PR 18-JUN-1999; 99US-0139763.  
PR 21-JUN-1999; 99US-0139817.  
PR 22-JUN-1999; 99US-0139899.  
PR 23-JUN-1999; 99US-0140353.  
PR 23-JUN-1999; 99US-0140354.  
PR 24-JUN-1999; 99US-0140685.  
PR 28-JUN-1999; 99US-0140823.  
PR 29-JUN-1999; 99US-0140991.  
PR 30-JUN-1999; 99US-0141287.  
PR 01-JUL-1999; 99US-0141842.  
PR 01-JUL-1999; 99US-0142154.  
PR 02-JUL-1999; 99US-0142055.  
PR 06-JUL-1999; 99US-0142390.  
PR 08-JUL-1999; 99US-0142803.  
PR 09-JUL-1999; 99US-0142920.  
PR 12-JUL-1999; 99US-0142977.

PR 13-JUL-1999; 99US-0143542.  
PR 14-JUL-1999; 99US-0143624.  
PR 15-JUL-1999; 99US-0144005.  
PR 16-JUL-1999; 99US-0144085.  
PR 19-JUL-1999; 99US-0144086.  
PR 19-JUL-1999; 99US-0144325.  
PR 19-JUL-1999; 99US-0144331.  
PR 19-JUL-1999; 99US-0144332.  
PR 19-JUL-1999; 99US-0144333.  
PR 19-JUL-1999; 99US-0144334.  
PR 19-JUL-1999; 99US-0144335.  
PR 20-JUL-1999; 99US-0144352.  
PR 20-JUL-1999; 99US-0144632.  
PR 20-JUL-1999; 99US-0144884.  
PR 21-JUL-1999; 99US-0144814.  
PR 21-JUL-1999; 99US-0145086.  
PR 21-JUL-1999; 99US-0145088.  
PR 22-JUL-1999; 99US-0145085.  
PR 22-JUL-1999; 99US-0145087.  
PR 22-JUL-1999; 99US-0145089.  
PR 22-JUL-1999; 99US-0145192.  
PR 23-JUL-1999; 99US-0145145.  
PR 23-JUL-1999; 99US-0145218.  
PR 23-JUL-1999; 99US-0145224.  
PR 26-JUL-1999; 99US-0145276.  
PR 27-JUL-1999; 99US-0145913.  
PR 27-JUL-1999; 99US-0145918.  
PR 27-JUL-1999; 99US-0145919.  
PR 28-JUL-1999; 99US-0145951.  
PR 02-AUG-1999; 99US-0146386.  
PR 02-AUG-1999; 99US-0146388.  
PR 03-AUG-1999; 99US-0146389.  
PR 03-AUG-1999; 99US-0147038.  
PR 04-AUG-1999; 99US-0147204.  
PR 04-AUG-1999; 99US-0147302.  
PR 05-AUG-1999; 99US-0147192.  
PR 05-AUG-1999; 99US-0147260.  
PR 06-AUG-1999; 99US-0147303.  
PR 06-AUG-1999; 99US-0147416.  
PR 09-AUG-1999; 99US-0147493.  
PR 09-AUG-1999; 99US-0147935.  
PR 10-AUG-1999; 99US-0148171.  
PR 11-AUG-1999; 99US-0148319.  
PR 12-AUG-1999; 99US-0148341.  
PR 13-AUG-1999; 99US-0148565.  
PR 13-AUG-1999; 99US-0148684.  
PR 16-AUG-1999; 99US-0149368.  
PR 17-AUG-1999; 99US-0149175.  
PR 18-AUG-1999; 99US-0149426.  
PR 20-AUG-1999; 99US-0149722.  
PR 20-AUG-1999; 99US-0149723.  
PR 20-AUG-1999; 99US-0149929.  
PR 23-AUG-1999; 99US-0149902.  
PR 23-AUG-1999; 99US-0149930.  
PR 25-AUG-1999; 99US-0150566.  
PR 26-AUG-1999; 99US-0150884.  
PR 27-AUG-1999; 99US-0151065.  
PR 27-AUG-1999; 99US-0151066.  
PR 27-AUG-1999; 99US-0151080.  
PR 30-AUG-1999; 99US-0151303.  
PR 31-AUG-1999; 99US-0151438.  
PR 01-SEP-1999; 99US-0151930.  
PR 07-SEP-1999; 99US-0152363.  
PR 10-SEP-1999; 99US-0153070.  
PR 13-SEP-1999; 99US-0153758.  
PR 15-SEP-1999; 99US-0154018.  
PR 16-SEP-1999; 99US-0154039.  
PR 20-SEP-1999; 99US-0154779.  
PR 22-SEP-1999; 99US-0155139.  
PR 23-SEP-1999; 99US-0155486.  
PR 24-SEP-1999; 99US-0155659.  
PR 28-SEP-1999; 99US-0156458.  
PR 29-SEP-1999; 99US-0156596.

PR 04-OCT-1999; 99US-0157117.  
PR 05-OCT-1999; 99US-0157753.  
PR 06-OCT-1999; 99US-0157865.  
PR 07-OCT-1999; 99US-0158029.  
PR 08-OCT-1999; 99US-0158232.  
PR 12-OCT-1999; 99US-0158359.  
PR 13-OCT-1999; 99US-0159293.  
PR 13-OCT-1999; 99US-0159294.  
PR 13-OCT-1999; 99US-0159295.  
PR 14-OCT-1999; 99US-0159329.  
PR 14-OCT-1999; 99US-0159330.  
PR 14-OCT-1999; 99US-0159331.  
PR 14-OCT-1999; 99US-0159637.  
PR 18-OCT-1999; 99US-0159638.  
PR 18-OCT-1999; 99US-0159584.  
PR 21-OCT-1999; 99US-0160741.  
PR 21-OCT-1999; 99US-0160767.  
PR 21-OCT-1999; 99US-0160768.  
PR 21-OCT-1999; 99US-0160770.  
PR 21-OCT-1999; 99US-0160814.  
PR 21-OCT-1999; 99US-0160815.  
PR 22-OCT-1999; 99US-0160980.  
PR 22-OCT-1999; 99US-0160981.  
PR 22-OCT-1999; 99US-0160989.  
PR 25-OCT-1999; 99US-0161404.  
PR 25-OCT-1999; 99US-0161405.  
PR 25-OCT-1999; 99US-0161406.  
PR 26-OCT-1999; 99US-0161359.  
PR 26-OCT-1999; 99US-0161360.  
PR 26-OCT-1999; 99US-0161361.  
PR 28-OCT-1999; 99US-0161920.  
PR 28-OCT-1999; 99US-0161992.  
PR 28-OCT-1999; 99US-0161993.  
PR 29-OCT-1999; 99US-0162142.

Alignment Scores:  
Pred. No.: 282  
Score: 45.00  
Percent Similarity: 72.73%  
Best Local Similarity: 72.73%  
Query Match: 50.00%  
DB: 21  
Gaps: 0  
Length: 2154  
Matches: 8  
Conservative: 0  
Mismatches: 3  
Indels: 0

US-10-020-441-3 (1-16) x AAC48833 (1-2154)  
CY 4 GYARGASPAASuASpARgASpProSeRg 14  
DB 370 GGATTGGATGACAGACGATCCGAGAGA 402

Search completed: April 8, 2003, 17:14:02  
Job time : 39 secs



GenCore version 5.1.4-p5\_4378  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - protein search, using frame\_plus\_n2p model

Run on: April 10, 2003, 07:47:30 ; Search time 60.5 Seconds  
(without alignments)  
4693.902 Million cell updates/sec

Title: US-10-020-441-1

Perfect score: 477

Sequence: 1 ggtgcaactcgtgtgatcacg.....ttgtacgcctcaatttga 1477

Scoring table:

OLIGO  
Xgapop 60.0 , Xgapext 60.0  
Ygapop 60.0 , Ygapext 60.0  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 283224 segs, 96134422 residues

Word size: 1

Total number of hits satisfying chosen parameters: 565536

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Listing first 1000 summaries

Command line parameters:

-MODEL=frame\_n2p.model -DEV=xlp  
-O=/cgn2.1/USPTO\_pool/US10020441/runat\_10042003\_074720\_25922/app\_query.fasta\_1.1671  
-DB=PIR\_73 -QFMT=fastan -SUFFIX=oligo.rpr -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0  
-UNITS=bits -START=1 -END=1 -MATRIX=oligo -TRANS=human40.cdi -LIST=1000  
-LOCALIGN=200 -THR\_SCORE=quality -THR\_MIN=1 -ALIGN=40 -MODE=LOCAL -OUTFMT=pro  
-NORMEXT -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000  
-USER=US10020441.ecgn.1.1.80.etrnat\_10042003\_074720\_25922 -NCPU=6 -ICPU=3  
-NO\_XLPHY -NO\_MAP -LARGEQUERY -NEG\_SCORES=0 -WAIT -LONGLOG -DEV\_TIMEOUT=120  
-WARN\_TIMEOUT=30 -THREADS=1 -XGAPOP=60 -XGAPEXT=60 -Fgapop=6 -Fgapext=7  
-YGAPOP=60 -YGAPEXT=60 -DELop=6 -DELEXt=7

Database :

PIR\_73:.\*  
1: PIR1:.\*  
2: PIR2:.\*  
3: PIR3:.\*  
4: PIR4:.\*

Pred. NO. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	106	22.2	264	2 A28942	pancreatic elastas
2	106	22.1	377	2 C88710	hypothetical prote
3	106	1.9	110	2 AB2370	hypothetical prote
4	9	1.9	382	2 B90786	hypothetical prote
5	9	1.9	382	2 H85645	hypothetical prote
6	9	1.9	382	2 B64843	hypothetical prote
7	9	1.9	696	2 D81346	probable iron upla
8	8	1.7	132	2 T28218	hypothetical prote
9	8	1.7	156	2 A97841	hypothetical prote
10	8	1.7	165	2 H84997	hypothetical prote
11	8	1.7	177	2 S23505	chymase (EC 3.4.21
12	8	1.7	219	2 B69425	conserved hypothet
13	8	1.7	226	2 JE0151	myonase (EC 3.4.21
14	8	1.7	232	2 S32398	serine proteinase

15	8	1.7	244	2 S26042	chymase (EC 3.4.21
16	8	1.7	246	1 A46504	chymase (EC 3.4.21
17	8	1.7	246	2 B38678	mast cell proteina
18	8	1.7	246	2 B38678	mast cell proteina
19	8	1.7	247	1 KYHUCM	chymase (EC 3.4.21
20	8	1.7	247	1 PRRG	mast cell proteina
21	8	1.7	247	2 S59135	mast cell proteina
22	8	1.7	247	2 S23504	chymase (EC 3.4.21
23	8	1.7	247	2 S64708	chymase (EC 3.4.21
24	8	1.7	249	1 A35842	chymase (EC 3.4.21
25	8	1.7	255	2 A27122	calthepin G (EC 3.
26	8	1.7	259	2 B31136	tissue kallikrein
27	8	1.7	260	2 S26043	chymase (EC 3.4.21
28	8	1.7	261	2 A31136	tissue kallikrein
29	8	1.7	278	2 AH0282	probable peptidase
30	8	1.7	281	2 T13596	trypsin homology -
31	8	1.7	299	1 PRSMBG	streptogrisin B (E
32	8	1.7	338	2 T45941	translation elonga
33	8	1.7	395	2 T31578	hypothetical prote
34	8	1.7	396	2 S56496	prophage p4 integr
35	8	1.7	421	2 C91109	integrase (impor
36	8	1.7	421	2 F85954	probable pathogeni
37	8	1.7	436	2 T36706	hypothetical prote
38	8	1.7	502	2 S61935	SKS1 protein - yea
39	8	1.7	563	2 S70196	kf1B protein - Esc
40	8	1.7	631	2 S21495	tomato leucine zip
41	8	1.7	747	2 T51579	complement factor
42	8	1.7	822	2 A38420	antifreeze glycop
43	8	1.7	891	2 T19915	hypothetical prote
44	8	1.7	1004	2 T30338	oviductin (EC 3.4.
45	8	1.7	1170	2 T45914	integrin alpha 2 s
46	7	1.5	14	2 S29209	avenin alpha-2 - o
47	7	1.5	52	2 S06211	hypothetical prote
48	7	1.5	52	2 B71281	hypothetical prote
49	7	1.5	61	2 PS0049	serine proteinase
50	7	1.5	66	2 T52972	kallikrein - mouse
51	7	1.5	73	2 S44462	elastase (EC 3.4.2
52	7	1.5	75	2 A37002	catrooxobin I (EC 3
53	7	1.5	77	2 T48725	Q300 protein - mu
54	7	1.5	83	2 T23366	hypothetical prote
55	7	1.5	83	2 AD0277	probable phage-rel
56	7	1.5	85	2 S44461	elastase (EC 3.4.2
57	7	1.5	93	2 PL0223	T-cell receptor be
58	7	1.5	93	2 PL0224	T-cell receptor be
59	7	1.5	96	2 A05308	tissue kallikrein
60	7	1.5	98	1 B0RT2	prostatic steroid-
61	7	1.5	102	2 S69877	hypothetical prote
62	7	1.5	104	2 S15395	tissue kallikrein-
63	7	1.5	113	2 E27664	T-cell receptor be
64	7	1.5	116	2 G72686	hypothetical prote
65	7	1.5	117	2 H64094	ribosomal protein
66	7	1.5	118	1 R3EC13	ribosomal protein
67	7	1.5	118	2 C91149	30S ribosomal subu
68	7	1.5	118	2 G83113	30S ribosomal prot
69	7	1.5	118	2 F832714	30S ribosomal prot
70	7	1.5	118	2 D83061	30S ribosomal prot
71	7	1.5	118	2 G85994	30S ribosomal subu
72	7	1.5	118	2 AH0028	30S ribosomal prot
73	7	1.5	118	2 AG1008	30S ribosomal chai
74	7	1.5	120	2 S05791	matng pheromone a
75	7	1.5	120	2 E81229	30S ribosomal prot
76	7	1.5	121	2 H71351	probable ribosomal
77	7	1.5	121	2 F81254	30S ribosomal prot
78	7	1.5	122	2 G87406	30S ribosomal protei
79	7	1.5	122	2 AC2813	SSU ribosomal prot
80	7	1.5	122	2 AE3349	30S ribosomal prot
81	7	1.5	122	2 AE2181	hypothetical prote
82	7	1.5	123	2 D97282	ribosomal protein
83	7	1.5	124	2 A70566	probable ribosomal
84	7	1.5	124	2 S62819	ribosomal protein
85	7	1.5	124	2 D64219	ribosomal protein
86	7	1.5	124	2 C87154	30S ribosomal prot
87	7	1.5	125	2 C70162	ribosomal protein

151  
136  
15













```

c 964 6 1.3 141 2 T33212 hypothetical prote
965 6 1.3 141 2 E64949 probable membrane
966 6 1.3 141 2 C90951 hypothetical prote
967 6 1.3 141 2 G85799 hypothetical prote
c 968 6 1.3 141 2 G91001 hypothetical prote
c 969 6 1.3 142 2 C72053 ribosomal protein
c 970 6 1.3 142 2 F86568 l17 ribosomal prot
971 6 1.3 142 2 S70872 tapA protein precu
972 6 1.3 142 2 T18166 hypothetical prote
c 973 6 1.3 143 1 B69007 probable transcrip
974 6 1.3 143 2 E97523 ferridoxin, 2fe-2s
975 6 1.3 143 2 E86740 hypothetical prote
976 6 1.3 143 2 F86888 hypothetical prote
977 6 1.3 144 2 H86567 nucleoside-2'-p kin
c 978 6 1.3 144 2 G72056 nucleoside diphosp
c 979 6 1.3 144 2 T35511 hypothetical prote
c 980 6 1.3 144 2 AC1907 hypothetical prote
c 981 6 1.3 144 2 G84935 hypothetical prote
c 982 6 1.3 144 2 B83069 hypothetical prote
c 983 6 1.3 144 2 G98013 2-isopropylmalate
984 6 1.3 145 1 R3H015 ribosomal protein
985 6 1.3 145 1 R3H015 ribosomal protein
986 6 1.3 145 1 R3H015 ribosomal protein
987 6 1.3 145 2 A34823 ribosomal protein
988 6 1.3 145 2 B34823 ribosomal protein
c 989 6 1.3 145 2 T39797 hypothetical prote
990 6 1.3 145 2 T24969 hypothetical prote
991 6 1.3 145 2 T21691 hypothetical prote
992 6 1.3 145 2 H87351 hypothetical prote
c 993 6 1.3 146 2 E54403 hemoglobin C beta
c 994 6 1.3 146 2 S16371 hemoglobin C beta
c 995 6 1.3 146 2 D54403 hemoglobin C beta
c 996 6 1.3 147 2 F70232 hypothetical prote
997 6 1.3 147 2 A69495 conserved hypochet
c 998 6 1.3 147 2 A42435 leech antiplatelet
999 6 1.3 147 2 A99145 hypothetical prote
c1000 6 1.3 147 2 E84727 hypothetical prote

```

## ALIGNMENTS

```

RESULT 1
A28942
Pancreatic elastase (EC 3.4.21.36) precursor - fluke (Schistosoma mansoni)
C:Species: Schistosoma mansoni
C>Date: 30-Jun-1989 #sequence_revision 30-Jun-1989 #text_change 26-Aug-1999
C:Accession: A28942
R:Newport, G.R.; McKerrow, J.H.; Hedstrom, R.; Pettit, M.; McGarrigle, L.; Barr, P.J.;
J. Biol. Chem. 263, 13179-13184, 1988
A:Title: Cloning of the proteinase that facilitates infection by schistosome parasites.
A:Reference number: A28942; M0ID:86350818; PMID:3166457
A:Accession: A28942
A:Molecule type: mRNA
A:Residues: 1-264 <NDM>
A:Cross-references: GB:J03946; NID:q160947; PID:q160948
C:Superfamily: trypsin; trypsin homology
C:Keywords: hydrolase; serine proteinase
F;27-256/Domain: trypsin homology <TRY>

```

```

Alignment Scores:
Pred. No.: 7.3e-102 Length: 264
Score: 106.00 Matches: 106
Percent Similarity: 100.008 Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 22.22% Indels: 0
DB: 2 Gaps: 0

```

US-10-020-441-1 (1-1477) x A28942 (1-264)

```

OY 310 GTTTCATTTCTCACATGAGGAATGGAGCACACAGCATCATCAACCAACCTCTGGA 369
Db 79 ValSerPheutRrLeuAtrgAsnGLYAspInclngInglYleHSHISGInProSecrGLy 98

```

```

OY 370 GTTAAGTGGCAGCAGATATACGCTCTGTATGTGGCAGACAGAGACCAATC 429
Db 99 VallysValAlaprogLYTYTMePProserCysmetserrAlaargInArgpJro 118
OY 430 GCACAGACACATCATGATTTGCATTTGTAATGCTGGCTCAATATGTCACCTTA 489
Db 119 AlaGIntrrLeuSerGlyPheAspIleAlaIleValMeLleuAlaGInMeValAsnLeu 138
OY 490 CAGATGGATCAGATGATCATCTGCTCCACAGCATCGATATATCCGCCACCTGGAAT 549
Db 139 GInserGlyIleArgValIleSerLeuProGInProSerAspIleProProgLYThr 158
OY 550 GGTGTTTCATTTGTTGTTATGAGAGGATGATACGACCGTGATCCGTCACTGAAGAT 609
Db 159 GlyAlaPheIleValGLTYT-GlyArgspAspAsnAspArgAspProSerArgLysasn 178
OY 610 GGTGGAATTTGAGGAA 627
Db 179 GlycylIleLeuLysLys 184

```

```

RESULT 2
C88710
protein C43G2.5 [imported] - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C>Date: 10-May-2001 #sequence_revision 10-May-2001 #text_change 10-May-2001
C:Accession: C88710
R:anonymous, The C. elegans Sequencing Consortium.
Science 282, 2012-2018, 1998
A:Title: Genome sequence of the nematode C. elegans: a platform for investigating bio
A:Reference number: A75000; M0ID:99069613; PMID:9851916
A:Note: see websites genome.wustl.edu/gsc/C.elegans/ and www.sanger.ac.uk/projects/C-
A:Note: published errata appeared in Science 283, 35, 1999; Science 283, 2103, 1999;
A:Accession: C88710
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-377 <STO>
A:Cross-references: GB:chr_IV; PIDN:AB09110.1; PID:q1572759; GSPDB:GN00022; CESP:C43
A:Note: similar to peptidase family S1
C:Genetics:
A:Gene: C43G2.5
A:Map position: 4

```

```

Alignment Scores:
Pred. No.: 0.206 Length: 377
Score: 100.00 Matches: 10
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 2.10% Indels: 0
DB: 2 Gaps: 0

```

US-10-020-441-1 (1-1477) x C88710 (1-377)

```

OY 1322 GCTCCAGGTGACAGTGTGACACTCTCCTC 1351
Db 241 AlaProGlyAspSerGlyGlyProLeuLeu 250

```

## RESULT 3

AB2370 hypothetical protein alr4514 [imported] - Nostoc sp. (strain PCC 7120)

```

C:Species: Nostoc sp.
A:Note: Nostoc sp. strain PCC 7120 is a synonym of Anabaena sp. strain PCC 7120
C>Date: 14-Dec-2001 #sequence_revision 14-Dec-2001 #text_change 30-Jun-2002
C:Accession: AB2370
R:Kaneko, T.; Nakamura, Y.; Molk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.; Iriju
Nakazaki, N.; Shimpo, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Yasuda, M.; Tabata
DNA Res. 8, 205-213, 2001
A:Title: Complete Genomic Sequence of the Filamentous Nitrogen-Fixing Cyanobacterium
A:Reference number: AB1807; M0ID:21595285; PMID:11759640
A:Accession: AB2370
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-110 <KUR>
A:Cross-references: GB:BA000019; PIDN:BA076213.1; PID:q17133650; GSPDB:GN00179

```



A:Experimental source: strain PCC 7120  
C:Genetics:  
A:Gene: alr4514

## Alignment Scores:

Pred. No.:	2.61	Length:	110
Score:	9.00	Matches:	9
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	1.90%	Indels:	0
DB:	2	Gaps:	0

US-10-020-441-1 (1-1477) x AB2370 (1-110)

OY 743 ACACACCGATGCTAGTACAGACA 717

DB 68 ThrTrmTrmGlySerGlyArgGlnThr 76

## RESULT 4

hypothetical protein ECs1258 [imported] - Escherichia coli (strain O157:H7, substrain R1

C:Species: Escherichia coli

C:Date: 18-Jul-2001 #sequence\_revision 18-Jul-2001 #text\_change 18-Jul-2001

C:Accession: B90786

R:Hayashi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.G. gasavara, N.; Yasunaga, T.; Kuhara, S.; Shiba, T.; Hattori, M.; Shinagawa, H. DNA Res. 8, 11-22, 2001

A:Title: Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and genc

A:Reference number: A95629; MUID:21156231; PMID:11258796

A:Accession: B90786

A:Status: preliminary

A:Molecule type: DNA

A:Cross-references: GB:BA000007; PIDN:BA834681.1; PID:q13360718; GSPDB:GN00154

A:Experimental source: strain O157:H7, substrain RMD 050952

C:Genetics:

A:Gene: ECs1258

## Alignment Scores:

Pred. No.:	2.28	Length:	382
Score:	9.00	Matches:	9
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	1.89%	Indels:	0
DB:	2	Gaps:	0

US-10-020-441-1 (1-1477) x B90786 (1-382)

OY 1430 TATGCCAGTGTGGCTAGATGTTGAT 1456

DB 327 TyrAlaSerValAlaArgMetLeuasp 335

## RESULT 5

hypothetical protein Z1511 [imported] - Escherichia coli (strain O157:H7, substrain EDL9

C:Species: Escherichia coli

C:Date: 16-Feb-2001 #sequence\_revision 16-Feb-2001 #text\_change 14-Sep-2001

C:Accession: H85645

R:Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; Mayhew

iller, L.; Grobbeck, E.J.; Davis, N.W.; Lim, A.; Dimalanta, E.; Potamousis, K.; Apodaca,

Nature 409, 529-533, 2001

A:Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.

A:Reference number: A85480; MUID:21074935; PMID:11206551

A:Accession: H85645

A:Status: preliminary

A:Molecule type: DNA

A:Cross-references: GB:AE005174; NID:q12514369; PIDN:AA655628.1; GSPDB:GN00145; UWGP:Z15

A:Experimental source: strain O157:H7, substrain EDL933

C:Genetics:

A:Gene: Z1511

## Alignment Scores:

Pred. No.:	2.14	Length:	696
Score:	9.00	Matches:	9
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	1.90%	Indels:	0
DB:	2	Gaps:	0

Pred. No.:	2.28	Length:	382
Score:	9.00	Matches:	9
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	1.89%	Indels:	0
DB:	2	Gaps:	0

US-10-020-441-1 (1-1477) x H85645 (1-382)

OY 1430 TATGCCAGTGTGGCTAGATGTTGAT 1456

DB 327 TyrAlaSerValAlaArgMetLeuasp 335

## RESULT 6

hypothetical protein b1012 - Escherichia coli (strain K-12)

C:Species: Escherichia coli

C:Date: 12-Sep-1997 #sequence\_revision 17-Sep-1997 #text\_change 01-Mar-2002

C:Accession: B64843

R:Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.;

A.; Rose, D.J.; Mau, B.; Shao, Y.

Science 277, 1453-1462, 1997

A:Title: The complete genome sequence of Escherichia coli K-12.

A:Reference number: A64720; MUID:97426617; PMID:9278503

A:Accession: B64843

A:Status: nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-382 <BLAT>

A:Cross-references: GB:AE000202; GB:U00096; NID:q1787233; PIDN:AACT4097.1; PID:q17872

A:Experimental source: strain K-12, substrain MG1655

## Alignment Scores:

Pred. No.:	2.28	Length:	382
Score:	9.00	Matches:	9
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	1.89%	Indels:	0
DB:	2	Gaps:	0

US-10-020-441-1 (1-1477) x B64843 (1-382)

OY 1430 TATGCCAGTGTGGCTAGATGTTGAT 1456

DB 327 TyrAlaSerValAlaArgMetLeuasp 335

## RESULT 7

probable iron uptake protein Cj0755 [imported] - Campylobacter jejuni (strain NCTC 11

C:Species: Campylobacter jejuni

C:Date: 31-Mar-2000 #sequence\_revision 31-Mar-2000 #text\_change 03-Jun-2002

C:Accession: D81346

R:Parkhill, J.; Wren, B.W.; Mungall, K.; Kelsey, J.M.; Churcher, C.; Basham, D.; Chl

C.W.; Quail, M.; Rajandream, M.A.; Rutherford, K.M.; VanVleet, A.; Whitehead, S.; Bar

Nature 403, 665-668, 2000

A:Title: The genome sequence of the food-borne pathogen Campylobacter jejuni reveals

A:Reference number: AB1250; MUID:20150912; PMID:10688204

A:Accession: D81346

A:Status: preliminary

A:Molecule type: DNA

A:Cross-references: GB:AL139076; GB:AL111168; NID:q6968128; PIDN:CAB73021.1; PID:q696

A:Experimental source: serotype O2, strain NCTC 11168

C:Genetics:

A:Gene: cjaA

C:Superfamily: ferrienterochelin receptor; tonB-dependent receptor amino-terminal hom

## Alignment Scores:

Pred. No.:	2.14	Length:	696
Score:	9.00	Matches:	9
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	1.90%	Indels:	0
DB:	2	Gaps:	0

US-10-020-441-1 (1-1477) x D81346 (1-696)

OY 443 CTGAGTCTGTGCGATTGCTCTC 417  
|||||  
Db 6 LeuserValCysAlaIleGlyLeuLeu 14

## RESULT 8

T28218

hypothetical protein ORP57 - Melanoplus sanguinipes entomopoxvirus

C:Species: Melanoplus sanguinipes entomopoxvirus

C:Date: 21-Jan-2000 #sequence\_revision 21-Jan-2000 #text\_change 16-Feb-2001

C:Accession: T28218

R:Alonso, C.L.; Tulman, E.R.; Lu, Z.; Oma, E.; Kutish, G.F.; Rock, D.L.

J. Virol. 73, 533-552, 1999

A:Title: The genome of Melanoplus sanguinipes entomopoxvirus.

A:Reference number: 220484; MUID:99102612; PMID:9847359

A:Accession: T28218

A>Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-132 <AF0>

A:Cross-references: EMBL:AF063866; NID:94049647; PIDN:AC97821.1; PID:94049861

A:Experimental source: Isolate Tuscon

C:Genetics:

A:Note: MSY057

C:Superfamily: Melanoplus sanguinipes entomopoxvirus hypothetical protein 57

## Alignment Scores:

Pred. No.:	28.5	Length:	132
Score:	8.00	Matches:	8
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	1.68%	Indels:	0
DB:	2	Gaps:	0

US-10-020-441-1 (1-1477) x T28218 (1-132)

OY 669 CAGATGTCAGCAGTATTTCGTG 692  
|||||

Db 16 GlnMetSerValSerTyrSerVal 23

## RESULT 9

A97841

hypothetical protein RC1129 [imported] - Rickettsia conorii (strain Malish 7)

C:Species: Rickettsia conorii

C:Date: 30-Sep-2001 #sequence\_revision 30-Sep-2001 #text\_change 30-Sep-2001

C:Accession: A97841

R:Ogata, H.; Audic, S.; Renesto-Audiffren, P.; Fournier, P.E.; Barbe, V.; Samson, D.; R

Science 293, 2093-2098, 2001

A:Title: Mechanisms of Evolution in Rickettsia conorii and Rickettsia prowazekii.

A:Reference number: A97700; MUID:21442074; PMID:11557893

A:Accession: A97841

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-156 <KUR>

A:Cross-references: GB:AE006914; PIDN:AAL03667.1; PID:g15620254; GSPDB:GN00173

C:Genetics:

A:Gene: RC1129

## Alignment Scores:

Pred. No.:	28	Length:	156
Score:	8.00	Matches:	8
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	1.68%	Indels:	0
DB:	2	Gaps:	0

US-10-020-441-1 (1-1477) x A97841 (1-156)

OY 1451 TTGGATTTTGTAGCGCTCCAAATATT 1474  
|||||

Db 62 LeuaspheValArgSerAsnIle 69

RESULT 10  
H84997  
hypothetical protein kdtb [imported] - Buchnera sp. (strain APS)

C:Species: Buchnera sp.

C:Date: 02-Mar-2001 #sequence\_revision 02-Mar-2001 #text\_change 23-Mar-2001

C:Accession: H84997

R:Shigenobu, S.; Matanabe, H.; Hattori, M.; Sakaki, Y.; Ishikawa, H.

Nature 407, 81-86, 2000

A:Title: Genome sequence of the endocellular bacterial symbiont of aphids Buchnera sp

A:Reference number: A84930; MUID:20445173; PMID:10993077

A:Accession: H84997

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-165 <STO>

A:Cross-references: GB:AP000398; GSPDB:GN00144

A:Experimental source: strain APS

C:Genetics:

A:Gene: kdtb; BU583

C:Superfamily: Lipopolysaccharide core biosynthesis protein kdtb

## Alignment Scores:

Pred. No.:	27.9	Length:	165
Score:	8.00	Matches:	8
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	1.69%	Indels:	0
DB:	2	Gaps:	0

US-10-020-441-1 (1-1477) x H84997 (1-165)

OY 470 AGCATTCACATTCGATTCGAAT 447  
|||||

Db 31 SerIleThrIleAlaIleSerAsn 38

## RESULT 11

S23505

chymase (EC 3.4.21.39) 2 - mouse (fragment)

C:Species: Mus musculus (house mouse)

C:Date: 22-Nov-1993 #sequence\_revision 26-Jul-1996 #text\_change 18-Jul-1997

C:Accession: S23505

R:Chu, W.; Johnson, D.A.; Muslich, P.R.

Biochim. Biophys. Acta 1121, 83-87, 1992

A:Title: Molecular cloning and characterization of mouse mast cell chymases.

A:Reference number: S23504; MUID:92287966; PMID:1376147

A:Accession: S23505

A:Molecule type: mRNA

A:Residues: 1-177 <CHU>

A:Cross-references: EMBL:M68899

C:Superfamily: trypsin; trypsin homology

C:Keywords: hydrolase; mast cell; serine proteinase

F:1-170/Domain: trypsin homology (fragment) <TRI>

## Alignment Scores:

Pred. No.:	27.6	Length:	177
Score:	8.00	Matches:	8
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	1.68%	Indels:	0
DB:	2	Gaps:	0

US-10-020-441-1 (1-1477) x S23505 (1-177)

OY 1328 GGTGACAGTGTGAGCTCTCTC 1351  
|||||

Db 131 GlyaspSerGlyGlyProLeuLeu 138

## RESULT 12

B69425

conserved hypothetical protein AF1403 - Archaeoglobus fulgidus

C:Species: Archaeoglobus fulgidus

C:Date: 05-Dec-1997 #sequence\_revision 05-Dec-1997 #text\_change 29-Sep-1999

C:Accession: B69425

R:Klenk, H.P.; Clayton, R.A.; Tomb, J.F.; White, O.; Nelson, K.E.; Ketchum, K.A.; Dod

.: Fleischmann, R.D.; Quackenbush, J.; Lee, N.H.; Sutton, G.G.; Gill, S.; Kirkness, E.F.; Glodok, A.; Zhou, L.; Overbeek, R.; Gocayne, J.D.; Weidman, J.F.; McDonald, L. Nature 390, 364-370, 1997

A:Authors: Utterback, T.; Cotton, M.D.; Spriggs, T.; Attiach, P.; Kaine, B.P.; Sykes, S. Smith, H.O.; Woese, C.R.; Venter, J.C.

A:Title: The complete genome sequence of the hyperthermophilic, sulfate-reducing archaea

A:Reference number: A69250; MUID:98049343; PMID:9389475

A:Accession: B69425

A:Status: preliminary; nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-219 <KLE>

A:Cross-references: GB:AE001007; GB:AE000782; NID:g2689330; PIDN:AAB89840.1; PID:g264916

C:Superfamily: hypothetical protein MJ1458

## Alignment Scores:

Pred. No.:	27	Length:	219
Score:	8.00	Matches:	8
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	1.68%	Indels:	0
DB:	2	Gaps:	0

US-10-020-441-1 (1-1477) x B69425 (1-219)

OY 493 AGTGAATCAGATGATGATGCTG 516

Db 168 SerGylleatGvalIleSerIeu 175

## RESULT 13

JE0151

myonase (EC 3.4.21.3) - mouse

C:Species: Mus musculus (house mouse)

C&gt;Date: 13-Jun-1998 #sequence\_revision 10-Jul-1998 #text\_change 07-May-1999

C:Accession: JE0151

R:Horl, S.; Ohnami, S.; Horl, C.; Nohhara, K.

J. Biochem. 123, 650-658, 1998

A:Title: Purification and characterization of myonase from X-chromosome linked muscular

A:Reference number: JE0151; MUID:98207054; PMID:9538257

A:Accession: JE0151

A:Molecule type: mRNA

A:Residues: 1-226 &lt;HOR&gt;

A:Experimental source: muscle

C:Comment: This protein has chymotrypsin-like activities and hydrolyzes the amide bonds

C:Superfamily: trypsin; trypsin homology

C:Keywords: hydrolase

F:1-219/Domain: trypsin homology &lt;TRY&gt;

## Alignment Scores:

Pred. No.:	26.9	Length:	226
Score:	8.00	Matches:	8
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	1.68%	Indels:	0
DB:	2	Gaps:	0

US-10-020-441-1 (1-1477) x JE0151 (1-226)

OY 1328 GGTGACAGTGTGACCTCTCTC 1351

Db 180 GlyaspSerGlyGlyProIeuIeu 187

## RESULT 14

S32398

serine proteinase P-11C (EC 3.4.21.-) - silkworm

C:Species: Bombyx mori (silkworm)

C&gt;Date: 08-Dec-1993 #sequence\_revision 01-Sep-1995 #text\_change 08-Sep-1997

C:Accession: S32398

R:Sezaki, T.; Hishida, T.; Ichikawa, K.; Asari, S.

FEBS Lett. 320, 35-37, 1993

A:Title: Amino acid sequence of alkaliphilic serine protease from silkworm, Bombyx mori,

A:Reference number: S32398; MUID:93215816; PMID:8462672

A:Accession: S32398

A:Status: preliminary

A:Molecule type: protein

A:Residues: 1-232 <SAS>

C:Superfamily: trypsin; trypsin homology

C:Keywords: hydrolase; serine proteinase

F:1-228/Domain: trypsin homology <TRY>

## Alignment Scores:

Pred. No.:	26.8	Length:	232
Score:	8.00	Matches:	8
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	1.68%	Indels:	0
DB:	2	Gaps:	0

US-10-020-441-1 (1-1477) x S32398 (1-232)

OY 1328 GGTGACAGTGTGACCTCTCTC 1351

Db 187 GlyaspSerGlyGlyProIeuIeu 194

## RESULT 15

S26042

chymase (EC 3.4.21.39) 4 precursor - mouse (fragment)

N:Alternate names: mast cell proteinase MMCP-4

C:Species: Mus musculus (house mouse)

C&gt;Date: 19-Mar-1998 #sequence\_revision 19-Mar-1998 #text\_change 07-May-1999

C:Accession: S26042

R:Huang, R.; Blom, T.; Hellman, L.

Eur. J. Immunol. 21, 1611-1621, 1991

A:Title: Cloning and structural analysis of MMCP-1, MMCP-4 and MMCP-5, three mouse ma

A:Reference number: S26041; MUID:91285010; PMID:2060576

A:Accession: S26042

A:Molecule type: mRNA

A:Residues: 1-244 &lt;HUA&gt;

A:Cross-references: EMBL:X68804

A:Note: the authors translated the codon TAC for residue 177 as Trp

C:Superfamily: trypsin; trypsin homology

C:Keywords: hydrolase; serine proteinase

F:1-15/Domain: signal sequence (fragment) #status predicted &lt;SIG&gt;

F:16-17/Domain: activation peptide #status predicted &lt;ACT&gt;

F:18-244/Product: chymase 4 #status predicted &lt;MAT&gt;

F:18-236/Domain: trypsin homology &lt;TRY&gt;

F:62,106,199/Active site: His, Asp, Ser #status predicted

## Alignment Scores:

Pred. No.:	26.7	Length:	244
Score:	8.00	Matches:	8
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	1.68%	Indels:	0
DB:	2	Gaps:	0

US-10-020-441-1 (1-1477) x S26042 (1-244)

OY 1328 GGTGACAGTGTGACCTCTCTC 1351

Db 197 GlyaspSerGlyGlyProIeuIeu 204

## RESULT 16

A46504

chymase (EC 3.4.21.39) precursor - mouse

N:Alternate names: mast cell proteinase I

C:Species: Mus musculus (house mouse)

C&gt;Date: 18-Jun-1993 #sequence\_revision 27-Jun-1994 #text\_change 18-Jun-1999

C:Accession: A46504; S26041; A30105; S35607

R:Ghildyal, N.; McNeill, H.P.; Stechschulte, S.; Austen, K.F.; Silberstein, D.; Gurish

J. Immunol. 149, 2123-2129, 1992

A:Title: IL-10 induces transcription of the gene for mouse mast cell protease-1, a se

A:Reference number: A46504; MUID:92388686; PMID:1517575

A:Accession: A46504

A:Molecule type: DNA

A:Residues: 1-246 &lt;GHI&gt;

A:Cross-references: GB:S44609; NID:g255141; PIDN:AAB23194.1; PID:g255142



KYHUCM  
 chymase (EC 3.4.21.39) precursor [validated] - human  
 N:Alternate names: chymase II; mast cell proteinase I  
 C:Species: Homo sapiens (man)  
 C:Date: 30-Sep-1992 #sequence\_revision 30-Sep-1992 #text\_change 08-Dec-2000  
 C:Accession: A40967; A40967; A39861; S15906; S33247; S56063; A23686; A60772  
 R:Urata, H.; Kinoshita, A.; Perez, D.M.; Misono, K.S.; Bumpus, F.M.; Graham, R.M.; Husai  
 J. Biol. Chem. 266, 17173-17179, 1991  
 A:Title: Cloning of the gene and cDNA for human heart chymase.  
 A:Reference number: A40967; MUID:91373329; PMID:1894611  
 A:Accession: A40967  
 A:Molecule type: mRNA  
 A:Residues: 1-247 <URA>  
 A:Cross-references: GB:M69136; NID:g180539; PIDN:AA52019.1; PID:g180540  
 A:Experimental source: ventricular myocardium  
 A:Accession: B40967  
 A:Molecule type: DNA  
 A:Residues: 1-247 <URA>  
 A:Cross-references: GB:M69137; NID:g180543; PIDN:AA52021.1; PID:g180544  
 A:Note: Part of this sequence, including the amino end of the mature protein, was confir  
 R:Caughy, G.H.; Zerweck, E.H.; Vanderslice, P.  
 J. Biol. Chem. 266, 12936-12963, 1991  
 A:Title: Structure, chromosomal assignment, and deduced amino acid sequence of a human g  
 A:Reference number: A39861; MUID:91303311; PMID:2071582  
 A:Accession: A39861  
 A:Molecule type: DNA  
 A:Residues: 1-247 <CAU>  
 A:Cross-references: GB:M64269; NID:g180541; PIDN:AA52020.1; PID:g180542  
 A:Note: These authors suggest that mast cells are the source of heart chymase  
 R:Jenne, D.E.; Tschopp, J.  
 Biochem. J. 276, 567-568, 1991  
 A:Title: Angiotensin II-forming heart chymase is a mast-cell-specific enzyme.  
 A:Reference number: S15906; MUID:91264818; PMID:2049082  
 A:Accession: S15906  
 A:Status: nucleic acid sequence not shown  
 A:Molecule type: DNA  
 A:Residues: 26-60 <JEN>  
 A:Cross-references: EMBL:X59072; NID:g29949; PIDN:CAA1796.1; PID:g29950  
 R:Sukenaga, Y.; Kido, H.; Neki, A.; Enomoto, M.; Ishida, K.; Takagi, K.; Katunuma, N.  
 FEBS Lett. 323, 119-122, 1993  
 A:Title: Purification and molecular cloning of chymase from human tonsils.  
 A:Reference number: S33247; MUID:93265916; PMID:8495723  
 A:Accession: S33247  
 A:Molecule type: mRNA  
 A:Residues: 22-27, 'S', 29-247 <SUK>  
 A:Cross-references: GB:S61334; NID:g409010; PIDN:AA62628.1; PID:g409011  
 A:Accession: S56063  
 A:Molecule type: protein  
 A:Residues: 22-27, 'S', 29-42 <SU2>  
 R:Urata, H.; Kinoshita, A.; Misono, K.S.; Bumpus, F.M.; Husain, A.  
 J. Biol. Chem. 265, 22348-22357, 1990  
 A:Title: Identification of a highly specific chymase as the major angiotensin II-forming  
 A:Reference number: A23686; MUID:91093078; PMID:2266130  
 A:Accession: A23686  
 A:Molecule type: protein  
 R:Schlechter, N.M.; Iranl, A.M.A.; Sprows, J.L.; Abernethy, J.; Wintroub, B.; Schwartz, I  
 J. Immunol. 145, 2652-2661, 1990  
 A:Title: Identification of a cathepsin B-like proteinase in the MC-TC type of human mast  
 A:Reference number: A60772; MUID:91010743; PMID:2212656  
 A:Accession: A60772  
 A:Molecule type: protein  
 A:Residues: 22-27, 'S', 29-56 <SCH>  
 C:Comment: Functions of this chymotrypsin-like serine proteinase after mast cell degranu  
 he cleavage of angiotensin I to angiotensin II is accomplished by this enzyme rather th  
 C:Genetics:  
 A:Gene: GDB:CMAL  
 A:Cross-references: GDB:127603; OMIM:118940  
 A:Map position: 14q11.2-14q11.2  
 A:Introns: 20/1; 70/2; 115/3; 200/3  
 C:Superfamily: trypsin; trypsin homology  
 C:Keywords: glycoprotein; hydrolase; serine proteinase

F:1-19/Domain: signal sequence #status predicted <SIG>  
 F:20-21/Domain: propeptide #status predicted <PRO>  
 F:22-247/Product: chymase #status experimental <MAT>  
 F:22-240/Domain: trypsin homology <TRY>  
 F:66,110,203/Active site: His, Asp, Ser #status predicted  
 F:80/Binding site: carbonylate (Asn) (covalent) #status experimental  
 F:103/Binding site: carbohydrate (Asn) (covalent) #status predicted  
 Alignment Scores:  
 Pred. No.: 26.7 Length: 247  
 Score: 8.00 Matches: 8  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 1.68% Indels: 0  
 DB: 1 Gaps: 0  
 ;  
 US-10-020-441-1 (1-1477) x KYHUCM (1-247)  
 QY 1328 GGTGACAGTGTGACACTCTCTC 1351  
 |||||||||||||||||||||  
 DB 201 GLYASPERGLYGLYProleulu 208  
 RESULT 20  
 PRTG  
 mast cell proteinase II (EC 3.4.21.-) precursor - rat  
 N:Alternate names: group-specific proteinase  
 C:Species: Rattus norvegicus (Norway rat)  
 C:Date: 30-Nov-1980 #sequence\_revision 19-Oct-1995 #text\_change 18-Jun-1999  
 C:Accession: A29548; A00957  
 R:Benfey, P.N.; Ylin, F.H.; Leder, P.  
 J. Biol. Chem. 262, 5377-5384, 1987  
 A:Title: Cloning of the mast cell proteinase, RMCP II. Evidence for cell-specific expre  
 A:Reference number: A92636; MUID:87165980; PMID:3549719  
 A:Accession: A29548  
 A:Molecule type: DNA; mRNA  
 A:Residues: 1-247 <BEN>  
 A:Cross-references: GB:J02712; NID:g205335; PIDN:AA6284.1; PID:g205336  
 R:Woodbury, R.G.; Katunuma, N.; Kobayashi, K.; Titani, K.; Neurath, H.  
 Biochemistry 17, 811-819, 1978  
 A:Title: Covalent structure of a group-specific protease from rat small intestine.  
 A:Reference number: A00957; MUID:78124137; PMID:629933  
 A:Accession: A00957  
 A:Molecule type: protein  
 A:Residues: 21-237, 'R', 239-244 <WOO>  
 A:Experimental source: small intestine  
 C:Comment: This enzyme has chymotrypsin-like specificity toward small substrates. It  
 the gut and lung.  
 C:Genetics:  
 A:Introns: 19/1; 69/2; 124/3; 199/3  
 C:Superfamily: trypsin; trypsin homology  
 C:Keywords: hydrolase; mast cell; serine proteinase  
 F:1-20/Domain: signal sequence #status predicted <SIG>  
 F:21-247/Product: mast cell proteinase II #status predicted <MAT>  
 F:21-239/Domain: trypsin homology <TRY>  
 F:50-66,143-208,174-187/Disulfide bonds: #status predicted  
 F:65,109,202/Active site: His, Asp, Ser #status predicted  
 Alignment Scores:  
 Pred. No.: 26.7 Length: 247  
 Score: 8.00 Matches: 8  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 1.68% Indels: 0  
 DB: 1 Gaps: 0  
 US-10-020-441-1 (1-1477) x PRTG (1-247)  
 QY 1328 GGTGACAGTGTGACACTCTCTC 1351  
 |||||||||||||||||||||  
 DB 200 GLYASPERGLYGLYProleulu 207  
 RESULT 21  
 S59135

mast cell proteinase 3 (EC 3.4.21.-) precursor - rat  
C:Species: Rattus norvegicus (Norway rat)  
C>Date: 15-Feb-1996 #sequence,revision 01-Mar-1996 #text\_change 20-Jun-2000  
C:Accession: S59135  
R:Id: H.; Itoh, H.; Tomita, M.; Murakumo, Y.; Kobayashi, T.; Maruyama, H.; Oseada, Y.; N  
Biochem. J. 311, 675-680, 1995  
A:Title: Cloning of the cDNA encoding a novel rat mast-cell proteinase, mCP-3, and its  
A:Reference number: S59135; MUID:9603070; PMID:7487912  
A:Accession: S59135  
A>Status: preliminary  
A:Molecule type: mRNA  
A:Residues: 1-247 <ID>  
A:Cross-references: EMBL:D8495; NID:g559376; PIDN:BAA07507.1; PID:g559377  
C:Superfamily: trypsin; trypsin homology  
C:Keywords: hydrolase; serine proteinase  
F:22-240/Domain: trypsin homology <TRY>  
F:66,110,203/Active site: His, Asp, Ser #status predicted

Alignment Scores:  
Pred. No.: 26.7 Length: 247  
Score: 8.00 Matches: 8  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 1.68% Indels: 0  
DB: 2 Gaps: 0

US-10-020-441-1 (1-1477) x S59135 (1-247)

OY 1328 GGTGACAGTGTGGACCTTCCTC 1351  
|||||  
Db 201 GYASPSERGIGLYGYPROLEULEU 208

RESULT 22  
S23504  
chymase (EC 3.4.21.39) 1 precursor - mouse  
C:Species: Mus musculus (house mouse)  
C>Date: 22-Nov-1993 #sequence,revision 03-Aug-1995 #text\_change 22-Jun-1999  
C:Accession: S23504  
R:Chu, W.; Johnson, D.A.; Musich, P.R.  
Biochim. Biophys. Acta 1121, 83-87, 1992  
A:Title: Molecular cloning and characterization of mouse mast cell chymases.  
A:Reference number: S23504; MUID:92287966; PMID:1376147  
A:Accession: S23504  
A:Molecule type: mRNA  
A:Residues: 1-247 <CHU>  
A:Cross-references: EMBL:M6898; NID:g199028; PIDN:AAA9492.1; PID:g199029  
C:Superfamily: trypsin; trypsin homology  
C:Keywords: hydrolase; mast cell; serine proteinase  
F:1-21/Domain: signal sequence #status predicted <SIG>  
F:23-247/Product: chymase #status predicted <MAT>  
F:22-240/Domain: trypsin homology <TRY>  
F:66,110,203/Active site: His, Asp, Ser #status predicted

Alignment Scores:  
Pred. No.: 26.7 Length: 247  
Score: 8.00 Matches: 8  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 1.68% Indels: 0  
DB: 2 Gaps: 0

US-10-020-441-1 (1-1477) x S23504 (1-247)

OY 1328 GGTGACAGTGTGGACCTTCCTC 1351  
|||||  
Db 201 GYASPSERGIGLYGYPROLEULEU 208

RESULT 23  
S64708  
chymase (EC 3.4.21.39) 2 precursor - Mongolian jird  
N:Alternate names: mast cell protease-2 precursor  
C:Species: Meriones unguiculatus (Mongolian jird)  
C>Date: 28-Oct-1996 #sequence,revision 13-Mar-1997 #text\_change 20-Jun-2000

C:Accession: S64708  
R:Itoh, H.; Murakumo, Y.; Tomita, M.; Ide, H.; Kobayashi, T.; Maruyama, H.; Horii, Y.  
Biochem. J. 314, 923-929, 1996  
A:Title: Cloning of the cDNAs for mast-cell chymases from the jejunum of Mongolian ge  
t cells of mice and rats.  
A:Reference number: S64707; MUID:96177868; PMID:8615790  
A:Accession: S64708  
A>Status: preliminary  
A:Molecule type: mRNA  
A:Residues: 1-247 <IT>  
A:Cross-references: EMBL:D45174; NID:9633079; PIDN:BAA08122.1; PID:9633080  
C:Superfamily: trypsin; trypsin homology  
C:Keywords: hydrolase; serine proteinase  
F:1-19/Domain: signal sequence #status predicted <SIG>  
F:20-21/Domain: propeptide #status predicted <PRO>  
F:22-247/Product: chymase 2 #status predicted <MAT>  
F:22-240/Domain: trypsin homology <TRY>  
F:66,110,203/Active site: His, Asp, Ser #status predicted

Alignment Scores:  
Pred. No.: 26.7 Length: 247  
Score: 8.00 Matches: 8  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 1.68% Indels: 0  
DB: 2 Gaps: 0

US-10-020-441-1 (1-1477) x S64708 (1-247)

OY 1328 GGTGACAGTGTGGACCTTCCTC 1351  
|||||  
Db 201 GYASPSERGIGLYGYPROLEULEU 208

RESULT 24  
A35842  
chymase (EC 3.4.21.39) precursor - dog  
N:Alternate names: mast cell protease I; skeletal muscle (SK) protease  
C:Species: Canis lupus familiaris (dog)  
C>Date: 10-Sep-1999 #sequence,revision 10-Sep-1999 #text\_change 10-Sep-1999  
C:Accession: A35842; PT0062  
R:Caughy, G.H.; Raymond, W.W.; Vanderslice, P.  
Biochemistry 29, 5166-5171, 1990  
A:Title: Dog mast cell chymase: molecular cloning and characterization.  
A:Reference number: A35842; MUID:90335214; PMID:2378872  
A:Accession: A35842  
A:Molecule type: mRNA  
A:Residues: 1-249 <CAU>  
A:Cross-references: GB:J02904; NID:g163919; PIDN:AAA30835.1; PID:g163920  
A:Note: part of this sequence, including the amino end of the mature protein, was con  
R:Caughy, G.H.; Viro, N.F.; Lazarus, S.C.; Nadel, J.A.  
Biochim. Biophys. Acta 952, 142-149, 1988  
A:Title: Purification and characterization of dog mastocytoma chymase: identification  
A:Reference number: PT0062; MUID:88107816; PMID:3122835  
A:Accession: PT0062  
A:Molecule type: protein  
A:Residues: 22-28, 'R', 30-37, 'Q', 39-44, 'T', 46 <CA2>  
C:Superfamily: trypsin; trypsin homology  
C:Keywords: glycoprotein; hydrolase; serine proteinase; skeletal muscle  
F:1-21/Domain: signal sequence #status predicted <SIG>  
F:22-249/Product: chymase #status predicted <MAT>  
F:22-240/Domain: trypsin homology <TRY>  
F:66,110,203/Active site: His, Asp, Ser #status predicted  
F:121,155/Binding site: carbohydrate (Asn) (covalent) #status predicted

Alignment Scores:  
Pred. No.: 26.6 Length: 249  
Score: 8.00 Matches: 8  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 1.68% Indels: 0  
DB: 1 Gaps: 0

US-10-020-441-1 (1-1477) x A35842 (1-249)

OY 1328 GGTGACAGTGTGGACCTCTCCTC 1351  
 Db 201 GlyaspsergIglyProleu 208

RESULT 25

A27122  
 cathepsin G (EC 3.4.21.20) precursor - human  
 N:Alternate names: membrane-associated proteinase, U937 cell; neutrophil cathepsin G  
 C:Species: Homo sapiens (man)  
 C:Date: 05-Jun-1988 #sequence\_revision 05-Jun-1988 #text\_change 22-Jun-1999  
 C:Accession: A32627; A27122; A37115; S44427; A90031; A94428; A05307  
 R:Holm, P.A.; Popescu, N.C.; Hanson, R.D.; Salvesen, G.; Ley, T.J.  
 J. Biol. Chem. 264, 13412-13419, 1989  
 A:Title: Genomic organization and chromosomal localization of the human cathepsin G gene  
 A:Reference number: A32627; MUID:89340411; PMID:2569462  
 A:Accession: A32627  
 A:Molecule type: DNA  
 A:Residues: 1-255 <HCH>  
 A:Cross-references: GB:J04990; NID:g179914; PIDN:AAA51919.1; PID:g179915  
 R:Salvesen, G.; Parley, D.; Shuman, J.; Przydyla, A.; Reilly, C.; Travis, J.  
 Biochemistry 26, 2289-2293, 1987  
 A:Title: Molecular cloning of human cathepsin G: structural similarity to mast cell and  
 A:Reference number: A27122; MUID:87299663; PMID:3304423  
 A:Accession: A27122  
 A:Molecule type: mRNA  
 A:Residues: 1-255 <SAL>  
 A:Cross-references: GB:M6117; NID:g18181; PIDN:AAA52126.1; PID:g18182  
 R:Malison, C.M.; Villiers, C.L.; Colomb, M.G.  
 J. Immunol. 147, 921-926, 1991  
 A:Title: Proteolysis of C3 on U937 cell plasma membranes. Purification of cathepsin G.  
 A:Reference number: A46471; MUID:91318179; PMID:1861080  
 A:Accession: A46471  
 A>Status: preliminary  
 A:Molecule type: protein  
 A:Residues: 21-30 <MAI>  
 A:Experimental source: monocytic cell line U937  
 A:Note: sequence extracted from NCBI backbone (NCBI:44940)  
 R:Bangalore, N.; Travis, J.; Onunka, V.C.; Pohl, J.; Shafer, W.M.  
 J. Biol. Chem. 265, 13584-13588, 1990  
 A:Title: Identification of the primary antimicrobial domains in human neutrophil catheps  
 A:Reference number: A37115; MUID:90337964; PMID:2116408  
 A:Accession: A37115  
 A>Status: preliminary  
 A:Molecule type: protein  
 A:Residues: 21-25;97-103 <BAN>  
 R:Avril, L.E.; di Martino-Ferrier, M.; Pignede, G.; Seman, M.; Gauthier, F.  
 FEBS Lett. 345, 81-86, 1994  
 A:Title: Identification of the U-937 membrane-associated proteinase interacting with the  
 A:Reference number: S44427; MUID:94252410; PMID:8194606  
 A:Accession: S44427  
 A:Molecule type: protein  
 A:Residues: 21-52 <AVR>  
 A:Note: this protein was demonstrated at the surface of U-937 promonocytic cells; It cat  
 R:Heck, L.W.; Rostand, K.S.; Hunter, F.A.; Brown, A.  
 Anal. Biochem. 158, 217-227, 1986  
 A:Reference number: A90031; MUID:87097924; PMID:3799965  
 A:Accession: A90031  
 A:Molecule type: protein  
 A:Residues: 21-38; 'E', '40', 'T', '42', 'G', '44-45 <HEC>  
 A:Note: residue 41 was not identified  
 R:Travis, J.; Giles, P.J.; Porcelli, L.; Reilly, C.F.; Baugh, R.; Powers, J.  
 In Protein Degradation in Health and Disease, Ciba Foundation Symposium 75, 51-68, 1980  
 A:Reference number: A94428  
 A:Accession: A94428  
 A:Molecule type: protein  
 A:Residues: 21-30; 'R', '32-38', 'E', '40', 'T' <PRA>  
 C:Comment: This serine proteinase is found in the azurophil granules of neutrophils and  
 C:Genetics:  
 A:Gene: GDB:CTS9  
 A:Cross-references: GDB:119822; OMIM:116830  
 A:Map position: 14q11.2-14q11.2  
 A:Introns: 19/1; 68/2; 113/3; 198/3

C:Superfamily: trypsin; trypsin homology  
 C:Keywords: glycoprotein; hydrolase; lysosome; serine proteinase  
 F:1-18/Domain: signal sequence #status predicted <SIG>  
 F:19-20/Domain: propeptide #status predicted <AP>  
 F:21-255/Product: cathepsin G #status experimental <MAT>  
 F:21-238/Domain: trypsin homology <TRY>  
 F:49-65, 142-207, 172-186/Disulfide bonds: #status predicted  
 F:64, 108, 201/Active site: His, Asp, Ser #status predicted

Alignment Scores:

Pred. No.:	26.6	Length:	255
Score:	8.00	Matches:	8
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	1.68%	Indels:	0
DB:	2	Gaps:	0

US-10-020-441-1 (1-1477) x A27122 (1-255)

OY 1328 GGTGACAGTGTGGACCTCTCCTC 1351  
 Db 199 GlyaspsergIglyProleu 206

RESULT 26

B31136  
 tissue kallikrein (EC 3.4.21.35) 3 precursor, submandibular - rat  
 N:Alternate names: glandular prokallikrein 3, submandibular  
 C:Species: Rattus norvegicus (Norway rat)  
 C:Date: 31-Mar-1990 #sequence\_revision 31-Mar-1990 #text\_change 22-Jun-1999  
 C:Accession: B31136  
 R:Chen, Y.P.; Chao, J.; Chao, L.  
 Biochemistry 27, 7189-7196, 1988  
 A:Title: Molecular cloning and characterization of two rat renal kallikrein genes.  
 A:Reference number: A31136; MUID:89088074; PMID:2849988  
 A:Accession: B31136  
 A:Molecule type: DNA  
 A:Residues: 1-259 <CHB>  
 A:Cross-references: GB:M19648; GB:J02837; NID:g205002; PIDN:AAA51640.1; PID:g205004  
 A:Note: the authors translated the codon GTC for residue 230 as Cys  
 C:Superfamily: trypsin; trypsin homology  
 C:Keywords: hydrolase; serine proteinase  
 F:1-18/Domain: signal sequence #status predicted <SIG>  
 F:19-259/Product: tissue prokallikrein 3, submandibular #status predicted <MAT>  
 F:25-251/Domain: trypsin homology <TRY>  
 F:63, 118, 211/Active site: His, Asp, Ser #status predicted

Alignment Scores:

Pred. No.:	26.5	Length:	259
Score:	8.00	Matches:	8
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	1.68%	Indels:	0
DB:	2	Gaps:	0

US-10-020-441-1 (1-1477) x B31136 (1-259)

OY 1328 GGTGACAGTGTGGACCTCTCCTC 1351  
 Db 209 GlyaspsergIglyProleu 216

RESULT 27

S26043  
 clymase (EC 3.4.21.39) 5 precursor - mouse  
 N:Alternate names: mast cell proteinase MMCP-5; mast cell serine proteinase 5  
 C:Species: Mus musculus (house mouse)  
 C:Date: 13-Jan-1995 #sequence\_revision 13-Jan-1995 #text\_change 22-Jun-1999  
 C:Accession: S26043; A41076; B41076; C35646  
 R:Huang, R.; Blom, T.; Hellman, L.  
 Eur. J. Immunol. 21, 1611-1621, 1991  
 A:Title: Cloning and structural analysis of MMCP-1, MMCP-4 and MMCP-5, three mouse ma  
 A:Reference number: S26041; MUID:91285010; PMID:2060576  
 A:Accession: S26043  
 A:Molecule type: mRNA

A:Residues: 1-260 <HDA>  
A:Cross-references: EMBL:X68805; NID:953158; PIDN:CAA48705.1; PID:953159  
R:McNell, H.P.; Austen, K.F.; Somerville, L.L.; Gurish, M.F.; Stevens, R.L.  
J. Biol. Chem. 266, 20316-20322, 1991  
A:Title: Molecular cloning of the mouse mast cell protease-5 gene. A novel secretory gr  
A:Reference number: A41076; MUID:92041862; PMID:1939089  
A:Accession: A41076  
A:Molecule type: DNA  
A:Residues: 14-260 <MC1>  
A:Cross-references: GB:M73760  
A:Accession: B41076  
A:Molecule type: mRNA  
A:Residues: 14-260 <MC2>  
A:Cross-references: GB:M73759  
R:Keynolds, D.S.; Stevens, R.L.; Lane, W.S.; Carr, M.H.; Austen, K.F.; Serafin, W.E.  
Proc. Natl. Acad. Sci. U.S.A. 87, 3230-3234, 1990  
A:Title: Different mouse mast cell populations express various combinations of at least  
A:Reference number: A35646; MUID:90222202; PMID:2326280  
A:Accession: C35646  
A>Status: preliminary  
A:Molecule type: protein  
A:Residues: 35-40, 'X', 42-63, 'R' <REV>  
C:Genetics:  
A:Initons: 42/2; 92/3; 138/1; 223/1  
C:Superfamily: trypsin; trypsin homology  
C:Keywords: hydrolase; mast cell; serine proteinase; zymogen  
F:1-32/Domain: signal sequence #status predicted <Sig>  
F:33-34/Domain: activation peptide #status predicted <ACT>  
F:35-260/Product: chymase 5 #status experimental <MAT>  
F:35-253/Domain: trypsin homology <TRY>  
F:79,123,216/Active site: His, Asp, Ser #status predicted

Alignment Scores:  
Pred. No.: 26.5 Length: 260  
Score: 8.00 Matches: 8  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 1.68% Indels: 0  
DB: 2 Gaps: 0

US-10-020-441-1 (1-1477) x S26043 (1-260)  
QY 1328 GGTGACAGTGTGACCTCTCCTC 1351  
|||||  
Db 214 GLYAspSerGlyGlyProLeuLeu 221

RESULT 28  
A31136  
tissue kallikrein (EC 3.4.21.35) 7 precursor, submandibular - rat  
N:Alternate names: glandular prokallikrein 7, submandibular; proteinase A  
C:Species: Rattus norvegicus (Norway rat)  
C:date: 31-Mar-1990 #sequence\_revision 31-Mar-1990 #text\_change 22-Jun-1999  
C:Accession: A31136; S10698; S10699; D41429; B41429; S09315  
R:Chen, Y.P.; Chao, J.; Chao, L.  
Biochemistry 27, 7189-7196, 1988  
A:Title: Molecular cloning and characterization of two rat renal kallikrein genes.  
A:Reference number: A31136; MUID:89088074; PMID:2849988  
A:Accession: A31136  
A:Molecule type: DNA  
A:Residues: 1-261 <CHE>  
A:Cross-references: GB:M19647; GB:J02837; NID:9204999; PIDN:AAA41461.1; PID:9205000  
R:Elmoujahed, A.; Gultman, N.; Brillard, M.; Gauthier, F.  
FEBS Lett. 265, 137-140, 1990  
A:Title: Substrate specificity of two kallikrein family gene products isolated from the  
A:Reference number: S10698; MUID:90306305; PMID:2194829  
A:Accession: S10698  
A:Molecule type: protein  
A:Residues: 25-36 <ELM>  
A:Accession: S10699  
A:Molecule type: protein  
A:Residues: 112-139 <EL2>  
R:Kato, H.; Nakashima, E.; Enjoji, K.; Hayashi, I.; Oh-Ishii, S.; Iwanaga, S.  
J. Biochem. 102, 1389-1404, 1987

A:Title: Characterization of serine proteinases isolated from rat submaxillary gland:  
A:Reference number: A41429; MUID:88198057; PMID:3482210  
A:Accession: D41429  
A:Molecule type: protein  
A:Residues: 112-133 <KAT>  
A:Accession: B41429  
A:Molecule type: Protein  
A:Residues: 25-34, 'D', 36-45, 'S', 47-67, 'X', 69-75 <KA2>  
R:Brady, J.M.; MacDonald, R.J.  
Arch. Biochem. Biophys. 278, 342-349, 1990  
A:Title: The expression of two kallikrein gene family members in the rat kidney.  
A:Reference number: S09315; MUID:90225801; PMID:2183721  
A:Accession: S09315  
A>Status: not compared with conceptual translation  
A:Molecule type: mRNA  
A:Residues: 43-45, 'S', 47-114, 'A', 116-261 <BRA>  
C:Superfamily: trypsin; trypsin homology  
C:Keywords: hydrolase; serine proteinase  
F:1-18/Domain: signal sequence #status predicted <Sig>  
F:19-261/Product: tissue prokallikrein 7, submandibular #status predicted <MAT>  
F:25-253/Domain: trypsin homology <TRY>  
F:65,120,213/Active site: His, Asp, Ser #status predicted

Alignment Scores:  
Pred. No.: 26.5 Length: 261  
Score: 8.00 Matches: 8  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 1.68% Indels: 0  
DB: 2 Gaps: 0

US-10-020-441-1 (1-1477) x A31136 (1-261)  
QY 1328 GGTGACAGTGTGACCTCTCCTC 1351  
|||||  
Db 211 GLYAspSerGlyGlyProLeuLeu 218

RESULT 29  
AH0282  
probable pepetidase YP02319 [imported] - Yersinia pestis (strain CO92)  
C:Species: Yersinia pestis  
C:date: 02-Nov-2001 #sequence\_revision 02-Nov-2001 #text\_change 02-Nov-2001  
C:Accession: AH0282  
R:Parkhill, J.; Wren, B.W.; Thomson, N.R.; Titball, R.W.; Holden, M.T.G.; Prentice, M  
deno-Tarraga, A.M.; Chillingworth, T.; Cronin, A.; Davies, R.M.; Dougan, G  
ll, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.; Whitehead, S.; Barrel  
Nature 413, 523-527, 2001  
A:Title: Genome sequence of Yersinia pestis, the causative agent of plague.  
A:Reference number: AB0001; MUID:21470413; PMID:11586360  
A:Accession: AH0282  
A>Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-278 <YUR>  
A:Cross-references: GB:AL590842; PIDN:CAC91124.1; PID:915980316; GSPDB:GN00175  
C:Genetics:  
A:Gene: YP02319

Alignment Scores:  
Pred. No.: 26.3 Length: 278  
Score: 8.00 Matches: 8  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 1.68% Indels: 0  
DB: 2 Gaps: 0

US-10-020-441-1 (1-1477) x AH0282 (1-278)  
QY 121 GTACTCAGCTGTGTCATTGTGTT 144  
|||||  
Db 78 ValLeuThrAlaGlyHisCysVal 85

RESULT 30  
T13596



trypsin homolog - fruit fly (*Drosophila melanogaster*)  
 N:Alternate names: protein EG:80H7.1  
 C:Species: *Drosophila melanogaster*  
 C>Date: 13-Aug-1999 #sequence\_revision 13-Aug-1999 #text\_change 17-Nov-2000  
 C:Accession: T13596  
 R:Benos, P.  
 A:Submitted to the EMBL Data Library, April 1999  
 A:Description: Sequencing the distal X chromosome of *Drosophila melanogaster*.  
 A:Reference number: 217667  
 A:Accession: T13596  
 A:Status: preliminary; translated from GB/EMBL/DBJ  
 A:Molecule type: DNA  
 A:Residues: 1-281 <BEN>  
 A:Cross-references: EMBL:AL031027; PIDN:CAA19843.2  
 C:Genetics:  
 A:Cross-references: FlyBase:FBgn0000481  
 A:Introns: 159/3  
 A>Note: EG:80H7.1  
 C:Superfamily: trypsin; trypsin homology

Alignment Scores:  

Pred. No.:	26.3	Length:	281
Score:	8.00	Matches:	8
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	1.68%	Indels:	0
DB:	2	Gaps:	0

US-10-020-441-1 (1-1477) x T13596 (1-281)  
 QY 1328 GGTCAGAGTGTGACCTCTCCTC 1351  
 ||||||||||||||||||||  
 DB 208 GLYASPSERGLYGLYPROLEUEN 215

RESULT 31  
 PRSMBG  
 streptogrisin B (EC 3.4.21.81) precursor - *Streptomyces griseus*  
 C:Species: *Streptomyces griseus*  
 C>Date: 24-Apr-1984 #sequence\_revision 12-May-1994 #text\_change 16-Aug-2002  
 C:Accession: B26974; A00964  
 R:Henderson, G.; Krygsmann, P.; Liu, C.J.; Davey, C.C.; Malek, L.T.  
 J. Bacteriol. 169, 3778-3784, 1987  
 A:Title: Characterization and structure of genes for proteases A and B from *Streptomyces*  
 A:Reference number: A91842; MUID:87279934; PMID:3112129  
 A:Accession: B26974  
 A:Molecule type: DNA  
 A:Residues: 1-299 <HEN>  
 A:Cross-references: GB:M17104; NID:G153468; PIDN:AAA26819.1; PID:G153469  
 R:Jurasek, L.; Carpenter, M.R.; Smillie, L.B.; Gertler, A.; Levy, S.; Ericsson, L.H.  
 Biochem. Biophys. Res. Commun. 61, 1095-1100, 1974  
 A:Title: Amino acid sequence of *Streptomyces griseus* protease B, a major component of pr  
 A:Reference number: A00964; MUID:75127477; PMID:4218101  
 A:Accession: A00964  
 A:Molecule type: protein  
 A:Residues: 115-153, 'GT', 156, 'W', 158-203, 'B', 205-292, 'V', 294-295, 'A', 297-299 <JUR>  
 A:Experimental source: strain K-1  
 A>Note: disulfide bonds were determined  
 R:Debaere, L.T.J.; Hutcheon, W.L.B.; James, M.N.G.; Thiesen, W.E.  
 Nature 257, 758-763, 1975  
 A:Title: Tertiary structural differences between microbial serine proteases and pancreat  
 A:Reference number: A37556; MUID:76051298; PMID:1186854  
 A:Contents: annotation: X-ray crystallography, 2.8 angstroms  
 A>Note: on the basis of the crystallographic analysis, Trp-43 and Val-126 were added to  
 R:Debaere, L.T.J.; Brayer, G.D.; James, M.N.G.  
 Can. J. Biochem. 57, 135-144, 1979  
 A:Title: The 2.8-angstrom resolution structure of *Streptomyces griseus* protease b and 1t  
 A:Reference number: A37557; MUID:79210651; PMID:110426  
 A:Contents: annotation: X-ray crystallography, 2.8 angstroms  
 A>Note: on the basis of the crystallographic analysis, an alanine previously reported to  
 R:Gertler, A.  
 FEBS Lett. 43, 81-85, 1974  
 A:Title: Inhibition of *Streptomyces griseus* protease B by peptide chloromethyl ketones:  
 A:Reference number: A37558; MUID:74302902; PMID:421092

A:Contents: annotation: active site His  
 R:Wahby, S.; Engstrom, L.  
 Biochim. Biophys. Acta 151, 402-408, 1968  
 A:Title: Studies on *Streptomyces griseus* protease: II. The amino acid sequence around  
 A:Reference number: A37559; MUID:68124983; PMID:5636372  
 A:Contents: annotation: active site Ser  
 C:Genetics:  
 A:Gene: sprB  
 C:Superfamily: streptogrisin A  
 C:Keywords: hydrolase; serine proteinase  
 F:1-38/Domain: signal sequence #status predicted <SIG>  
 F:39-114/Domain: signal sequence #status predicted <PRO>  
 F:115-299/Product: proteinase B #status experimental <MAT>  
 F:128-148, 249-276/Disulfide bonds: #status experimental  
 F:147, 255/Active site: His, Ser #status experimental  
 F:177/Active site: Asp #status predicted

Alignment Scores:  

Pred. No.:	26.1	Length:	299
Score:	8.00	Matches:	8
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	1.68%	Indels:	0
DB:	1	Gaps:	0

US-10-020-441-1 (1-1477) x PRSMBG (1-299)  
 QY 1325 CCAGTGACAGTGTGACCTCTC 1348  
 ||||||||||||||||||||  
 DB 252 PROGLYASPSERGLYGLYPROLEU 259

RESULT 32  
 I45941  
 translation elongation factor EF-Ts - bovine  
 C:Species: *Bos primigenius taurus* (cattle)  
 C>Date: 16-Aug-1996 #sequence\_revision 16-Aug-1996 #text\_change 11-Jan-2000  
 C:Accession: I45941  
 R:Xin, H.; Morlaix, V.; Burkhart, W.; Spremull, L.L.  
 J. Biol. Chem. 270, 17243-17249, 1995  
 A:Title: Cloning and expression of mitochondrial translational elongation factor Ts f  
 A:Reference number: I45941; MUID:95340508; PMID:7615523  
 A:Accession: I45941  
 A:Status: preliminary; translated from GB/EMBL/DBJ  
 A:Molecule type: mRNA  
 A:Residues: 1-338 <XIN>  
 A:Cross-references: GB:L37935; NID:G625014; PIDN:AAA96807.1; PID:G625015  
 C:Superfamily: translation elongation factor EF-Ts

Alignment Scores:  

Pred. No.:	25.8	Length:	338
Score:	8.00	Matches:	8
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	1.68%	Indels:	0
DB:	2	Gaps:	0

US-10-020-441-1 (1-1477) x I45941 (1-338)  
 QY 1340 GGACCTCTCCATCCCTCA 1363  
 ||||||||||||||||||||  
 DB 26 GLYPROLEUENPROSERLEUEN 33

RESULT 33  
 T1578  
 hypothetical protein Y105C5A.u - *Caenorhabditis elegans*  
 C:Species: *Caenorhabditis elegans*  
 C>Date: 29-Oct-1999 #sequence\_revision 29-Oct-1999 #text\_change 29-Oct-1999  
 C:Accession: T1578  
 R:McMurray, A.  
 submitted to the EMBL Data Library, September 1999  
 A:Reference number: Z21045  
 A:Accession: T1578  
 A:Status: preliminary; translated from GB/EMBL/DBJ

A: Molecule type: DNA  
A: Residues: 1-395 <MIL>  
A: Cross-references: EMBL:AL117193; NID: e1549703; PIDN: CAB55001.1; CESP: Y105C5A.n  
A: Experimental source: clone Y105C5A  
C: Genetics:  
A: Gene: CESP: Y105C5A.n  
A: Introns: 59/3; 119/3; 231/3; 294/1; 339/2; 380/3

Alignment Scores:  
Pred. No.: 25.3 Length: 395  
Score: 8.00 Matches: 8  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 1.68% Indels: 0  
DB: 2 Gaps: 0

US-10-020-441-1 (1-1477) x T31578 (1-395)

QY 1029 TCTGCATCTCACCAGCTCTCCAAAC 1006  
|||||  
Db 138 SerProSerHisHisSerProAsn 145

RESULT 34  
S56496  
Prophage P4 integrase - Escherichia coli (strain K-12)  
N: Alternate names: hypothetical protein o396  
C: Species: Escherichia coli  
C: Date: 28-Oct-1995 #sequence\_revision 03-Nov-1995 #text\_change 01-Mar-2002  
C: Accession: S56496; A65240  
R: Burland, V.; Plunkett III, G.; Sofia, H.J.; Daniels, D.L.; Blattner, F.R.  
Nucleic Acids Res. 23, 2105-2119, 1995  
A: Title: Analysis of the Escherichia coli genome VI: DNA sequence of the region from 92.  
A: Reference number: S56314; MUID: 9534362; PMID: 7610040  
A: Accession: S56496  
A: Status: preliminary; nucleic acid sequence not shown; translation not shown  
A: Molecule type: DNA  
A: Residues: 1-396 <BUR>  
A: Cross-references: EMBL:U14003; NID: g1263172; PIDN: AAA97167.1; PID: g537112  
A: Note: The nucleotide sequence was submitted to the EMBL Data Library, August 1994  
R: Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.; Co  
A.; Rose, D.J.; Mau, B.; Shao, Y.  
Science 277, 1453-1462, 1997  
A: Title: The complete genome sequence of Escherichia coli K-12.  
A: Reference number: A64720; MUID: 97426617; PMID: 9278503  
A: Accession: A65240  
A: Status: preliminary; nucleic acid sequence not shown; translation not shown  
A: Molecule type: DNA  
A: Residues: 1-396 <BLAT>  
A: Cross-references: GB: AE000498; GB: U00096; NID: g2367368; PIDN: AAC77227.1; PID: g1790722.  
A: Experimental source: strain K-12, substrain MG1655  
C: Genetics:  
A: Gene: intB  
C: Superfamily: satellite phage P4 integrase

Alignment Scores:  
Pred. No.: 25.3 Length: 396  
Score: 8.00 Matches: 8  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 1.68% Indels: 0  
DB: 2 Gaps: 0

US-10-020-441-1 (1-1477) x S56496 (1-396)

QY 914 CACTTGAGAGAGAGACTCATG 937  
|||||  
Db 359 HistLeuGluGuaArgArgLeuMet 366

RESULT 35  
C91109  
Integrase [imported] - Escherichia coli (strain O157:H7, substrain RIMD 0509952)  
C: Species: Escherichia coli  
C: Date: 18-Jul-2001 #sequence\_revision 18-Jul-2001 #text\_change 03-Aug-2001

C: Accession: C91109  
R: Hayashi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C  
gasawara, N.; Yasunaga, T.; Kuhara, S.; Shiba, T.; Hattori, M.; Shinagawa, H.  
DNA Res. 8, 11-22, 2001  
A: Title: Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and g  
A: Reference number: A9629; MUID: 21156231; PMID: 11258796  
A: Accession: C91109  
A: Status: preliminary  
A: Molecule type: DNA  
A: Residues: 1-421 <HAY>  
A: Cross-references: GB: BA000007; PIDN: BAH37266.1; PID: g13363315; GSPDB: GN00154  
A: Experimental source: strain O157:H7, substrain RIMD 0509952  
C: Genetics:  
A: Gene: ECS3843  
C: Superfamily: satellite phage P4 integrase

Alignment Scores:  
Pred. No.: 25.2 Length: 421  
Score: 8.00 Matches: 8  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 1.68% Indels: 0  
DB: 2 Gaps: 0

US-10-020-441-1 (1-1477) x C91109 (1-421)

QY 914 CACTTGAGAGAGAGACTCATG 937  
|||||  
Db 384 HistLeuGluGuaArgArgLeuMet 391

RESULT 36  
F85954  
Probable pathogenicity island integrase 24313 [imported] - Escherichia coli (strain O  
C: Species: Escherichia coli  
C: Date: 16-Feb-2001 #sequence\_revision 16-Feb-2001 #text\_change 14-Sep-2001  
C: Accession: F85954  
R: Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; May  
Hiller, L.; Grothbeck, E.J.; Davis, N.W.; Lim, A.; Dimalanta, E.; Potamousis, K.; Apoda  
Nature 409, 529-533, 2001  
A: Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.  
A: Reference number: A85480; MUID: 21074935; PMID: 11206551  
A: Accession: F85954  
A: Status: preliminary  
A: Molecule type: DNA  
A: Residues: 1-421 <STO>  
A: Cross-references: GB: AE005174; NID: g12517519; PIDN: AAG58098.1; GSPDB: GN00145; UWGP:  
A: Experimental source: strain O157:H7, substrain EDL933  
C: Genetics:  
A: Gene: 24313  
C: Superfamily: satellite phage P4 integrase

Alignment Scores:  
Pred. No.: 25.2 Length: 421  
Score: 8.00 Matches: 8  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 1.68% Indels: 0  
DB: 2 Gaps: 0

US-10-020-441-1 (1-1477) x F85954 (1-421)

QY 914 CACTTGAGAGAGAGACTCATG 937  
|||||  
Db 384 HistLeuGluGuaArgArgLeuMet 391

RESULT 37  
T36706  
hypothetical protein SCH69.07c - Streptomyces coelicolor  
C: Species: Streptomyces coelicolor  
C: Date: 03-Dec-1999 #sequence\_revision 03-Dec-1999 #text\_change 03-Dec-1999  
C: Accession: T36706  
R: Murphy, L.; Harris, D.; Bentley, S.D.; Parkhill, J.; Barrell, B.G.; Rajandream, M.A  
submitted to the EMBL Data Library, June 1999

A:Reference number: Z21612  
A:Accession: T36706  
A:Title: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-436 <MUN>  
A:Cross-references: EMBL:AL079308; PIDN:CAB45204.1; GSPDB:GN00070; SCOEDB:SCH69.07c  
A:Experimental source: strain A3(2)  
C:Genetics:  
A:Gene: SCOEDB:SCH69.07c

Alignment Scores:  
Pred. No.: 25.1 Length: 436  
Score: 8.00 Matches: 8  
Percent Similarity: 100.00% Conservatave: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 1.68% Indels: 0  
DB: 2 Gaps: 0

US-10-020-441-1 (1-1477) x T36706 (1-436)

OY 280 TTAATGCTATATCTGCTCTTA 303  
|||||  
Db 135 LeuMetSerIleTyrLeuValIleu 142

RESULT 38  
S61935  
SKS1 protein - yeast (Saccharomyces cerevisiae)  
N:Alternate names: protein LPB5c; protein YPL026c; SHA3 protein  
C:Species: Saccharomyces cerevisiae  
C:Date: 10-Apr-1996 #sequence\_revision 19-Apr-1996 #text\_change 31-Mar-2000  
C:Accession: S61935; S63456  
R:Yang, Z.; Bisson, L.F.  
submitted to the EMBL Data Library, June 1995  
A:Reference number: S61935  
A:Accession: S61935  
A:Molecule type: DNA  
A:Residues: 1-502 <YAN>  
A:Cross-references: EMBL:U30613; NID:q1173540; PID:q1173541  
R:Mang, Y.; Ahmed, A.; Bussey, H.; Fortin, N.; Plesken, J.D.; Hall, J.; Storms, R.K.; VC  
submitted to the EMBL Data Library, September 1995  
A:Reference number: S63452  
A:Accession: S63456  
A:Molecule type: DNA  
A:Residues: 1-502 <MAN>  
A:Cross-references: EMBL:U36624; NID:q1276642; PID:q1039451; MIPS:YPL026c  
C:Genetics:  
A:Gene: SGD:SKS1; SHA3  
A:Cross-references: SGD:S0005947; MIPS:YPL026c  
A:Map position: 16L  
C:Superfamily: Saccharomyces cerevisiae SKS1 protein; protein kinase homology  
F:8-333/Domain: protein kinase homology #status atypical <KIN>

Alignment Scores:  
Pred. No.: 24.7 Length: 502  
Score: 8.00 Matches: 8  
Percent Similarity: 100.00% Conservatave: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 1.69% Indels: 0  
DB: 2 Gaps: 0

US-10-020-441-1 (1-1477) x S61935 (1-502)

OY 1223 ACAACACACACACACAGTACA 1200  
|||||  
Db 438 ThrAsnAsnHisAsnAsnSerThr 445

RESULT 39  
S70196  
KfIB protein - Escherichia coli  
C:Species: Escherichia coli  
C:Date: 12-Feb-1998 #sequence\_revision 20-Feb-1998 #text\_change 08-Dec-2000  
C:Accession: S70196; S45105  
R:Petit, C.; Rigby, G.P.; Pazzani, C.; Smith, A.; Sieberth, V.; Stevens, M.; Boulnois, G.

MOL. Microbiol. 17, 611-620, 1995  
A:Title: Region 2 of the Escherichia coli K5 capsule gene cluster encoding proteins f  
A:Reference number: S70191; MUID:96111482; PMID:8801416  
A:Accession: S70196  
A:Status: nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-563 <PET>  
A:Cross-references: EMBL:X77617; NID:q496601  
A:Experimental source: strain K5  
A:Note: The nucleotide sequence was submitted to the EMBL Data Library, February 1994  
A:Note: this translation is not annotated in Genbank entry ECKFAAF, release 117.0  
C:Genetics:  
A:Gene: KfIB

Alignment Scores:  
Pred. No.: 24.4 Length: 563  
Score: 8.00 Matches: 8  
Percent Similarity: 100.00% Conservatave: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 1.69% Indels: 0  
DB: 2 Gaps: 0

US-10-020-441-1 (1-1477) x S70196 (1-563)

OY 63 GATGAACGGGAATTCAGTGGGTG 40  
|||||  
Db 61 AspGluArgGluIleSerAlaVal 68

RESULT 40  
S21495  
tomato leucine zipper-containing protein - tomato  
C:Species: Lycopersicon esculentum (tomato)  
C:Date: 20-Feb-1995 #sequence\_revision 20-Feb-1995 #text\_change 24-Sep-1999  
C:Accession: S21495  
R:Gut, S.; Schmid, J.; Amrhein, N.  
submitted to the EMBL Data Library, May 1992  
A:Description: Deduced amino acid sequence of a plant cDNA containing a leucine zippe  
A:Reference number: S21495  
A:Accession: S21495  
A:Status: preliminary  
A:Molecule type: mRNA  
A:Residues: 1-631 <GUN>  
A:Cross-references: EMBL:Z12127; NID:q19274; PIDN:CAA78112.1; PID:q19275  
C:Superfamily: tomato leucine zipper-containing protein  
C:Keywords: leucine zipper

Alignment Scores:  
Pred. No.: 24.1 Length: 631  
Score: 8.00 Matches: 8  
Percent Similarity: 100.00% Conservatave: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 1.69% Indels: 0  
DB: 2 Gaps: 0

US-10-020-441-1 (1-1477) x S21495 (1-631)

OY 557 AAAACACACAGTTCAGAGTGGCGG 534  
|||||  
Db 381 LysThrProValProGlyGlyGly 388

Search completed: April 10, 2003, 07:58:34  
Job time : 99.5 secs

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.4 p5.4578  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 9, 2003, 12:52:47 ; Search time 46.0492 Seconds  
(without alignments)  
221.291 Million cell updates/sec

Title: US-10-020-441-2

Perfect score: 552  
Sequence: 1 VSFLTRNGDQGIIHQPSG.....YGRDNDPRPSRKNGILKK 106

Scoring table: BIOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database : PIR\_73:\*  
1: pirl:\*  
2: pirl2:\*  
3: pirl3:\*  
4: pirl4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	552	100.0	264	2 A28942	pancreatic elastase
2	108.5	19.7	237	2 S68702	trypsin (EC 3.4.21.1)
3	102	18.5	230	2 I48685	maat cell proteinase
4	101	18.5	276	2 A38654	maat cell proteinase
5	101	18.3	270	2 S6160	maat cell trypsin
6	101	18.3	274	2 A45754	trypsin (EC 3.4.21.1)
7	99	17.9	275	2 A35863	trypsin (EC 3.4.21.1)
8	99	17.9	275	2 C35863	trypsin (EC 3.4.21.1)
9	96	17.4	177	2 S23505	chymase (EC 3.4.21.1)
10	95.5	17.3	270	2 A29934	pancreatic elastase
11	93	16.8	246	2 A38678	maat cell proteinase
12	93	16.8	275	2 B35863	trypsin (EC 3.4.21.1)
13	92.5	16.8	270	2 B29934	pancreatic elastase
14	89.5	16.2	244	2 A34910	maat cell proteinase
15	89	16.1	274	2 J44171	trypsin (EC 3.4.21.1)
16	88.5	16.0	244	2 A46721	chymase (EC 3.4.21.1)
17	88.5	16.0	281	2 T13596	trypsin homolog -
18	85.5	15.5	269	2 B32410	maat cell proteinase
19	85	15.4	266	2 E4RT1	pancreatic elastase
20	83	15.0	273	2 A47246	trypsin (EC 3.4.21.1)
21	81.5	14.8	247	1 PRRTG	maat cell proteinase
22	81	14.7	1113	2 J50315	low-density lipoprotein
23	80	14.5	268	2 J01473	pancreatic elastase
24	80	14.5	711	1 A47136	macrophage-stimulating
25	80	14.5	1524	2 T30337	polypeptide - Afri
26	78.5	14.2	236	2 A28566	T-cell suppressor
27	78.5	14.2	258	2 S70439	pancreatic elastase
28	78.5	14.2	267	4 A56615	probable pancreatic
29	77.5	14.0	240	1 CPBOA1	procarboxypeptidase

30	77.5	14.0	271	1 E4RT2	pancreatic elastase
31	76	13.8	716	1 A40332	macrophage-stimulating
32	75	13.6	257	2 B45061	granzyme A (EC 3.4.21.1)
33	75	13.6	260	2 A45061	granzyme A (EC 3.4.21.1)
34	75	13.6	268	2 S68826	pancreatic elastase
35	75	13.6	268	2 S68825	pancreatic elastase
36	74.5	13.5	246	1 A46504	chymase (EC 3.4.21.1)
37	74.5	13.5	246	2 S64707	chymase (EC 3.4.21.1)
38	74.5	13.5	266	1 E4PG	pancreatic elastase
39	74.5	13.5	1121	2 T52631	1-phosphatidylinositol
40	74	13.4	246	2 B25528	trypsin (EC 3.4.21.1)
41	73.5	13.3	271	2 A25528	pancreatic elastase
42	73	13.2	256	1 PRRTG	proteinase 3 (EC 3.4.21.1)
43	72.5	13.1	269	2 T1265	aldehyde dehydrogenase
44	72	13.0	261	2 C26823	pancreatic elastase
45	72	13.0	271	2 S29239	chymotrypsin (EC 3.4.21.1)

## ALIGNMENTS

RESULT 1  
A28942  
pancreatic elastase (EC 3.4.21.36) precursor - fluke (Schistosoma mansoni)  
C:Species: Schistosoma mansoni  
C>Date: 30-Jun-1989 #sequence\_revision 30-Jun-1989 #text\_change 26-Aug-1999  
C:Accession: A28942  
R:Newport, G.R.; McKerrow, J.H.; Hesterrow, R.; Pettit, M.; McGarrigle, L.; Barr, P.J.;  
J. Biol. Chem. 263, 13179-13184, 1988  
A:Title: Cloning of the proteinase that facilitates infection by schistosome parasites  
A:Reference number: A28942; MUID:88330818; PMID:3166457  
A:Accession: A28942  
A:Molecule type: mRNA  
A:Residues: 1-264 <NEW>  
A:Cross-references: GB:U03946; NID:G160947; PID:G160948  
C:Superfamily: trypsin; trypsin homology  
C:Keywords: hydrolase; serine proteinase  
F:27-256/Domain: trypsin homology <TRY>

Query Match 100.0%; Score 552; DB 2; Length 264;  
Best Local Similarity 100.0%; Pred. No. 5.4e-53;  
Matches 106; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 VSFLTRNGDQGGIIHQPSGVKAPGKMPSCMSARORRPIAQTSGFDIAIYMLAQMNL 60  
Db 79 VSFLTRNGDQGGIIHQPSGVKAPGKMPSCMSARORRPIAQTSGFDIAIYMLAQMNL 138

Qy 61 QSGIRVLSLPQSDIPPGTGVFIYGRDNDPRPSRKNGILKK 106  
Db 139 QSGIRVLSLPQSDIPPGTGVFIYGRDNDPRPSRKNGILKK 184

RESULT 2  
S68702  
trypsin (EC 3.4.21.59) - bovine (fragment)  
C:Species: Bos primigenius taurus (cattle)  
C>Date: 25-Feb-1998 #sequence\_revision 13-Mar-1998 #text\_change 20-Jun-2000  
C:Accession: S68702; S68676  
R:Pallao, M.; Gambacorta, A.; Fiorucci, L.; Mignogna, G.; Barra, D.; Ascoti, F.  
Eur. J. Biochem. 237, 100-105, 1996  
A:Title: cDNA cloning and primary structure of trypsin from bovine mast cells, and evi-  
A:Reference number: S68676; MUID:96203914; PMID:8620861  
A:Accession: S68702  
A:Molecule type: mRNA  
A:Residues: 1-237 <PAL>  
A:Cross-references: EMBL:X94982; NID:G1332446; PID:CAA64438.1; PID:G1332447  
A:Accession: S68676  
A:Molecule type: Protein  
A:Residues: 1-8,61-74,90-97,126-148,162-190,208-222,227-235 <PAL>  
C:Superfamily: trypsin; trypsin homology  
C:Keywords: glycoprotein; hydrolase; serine proteinase; zymogen  
F:1-229/Domain: trypsin homology (fragment) <TRY>  
F:36,83,177/Active site: His, Asp, Ser #status predicted



A:Residues: 1-274 <M12>  
A:Cross-references: GB:M30038  
A:Note: the sequence from Fig. 4 is inconsistent with that from Fig. 2 in having 205-Pro  
C:Superfamily: trypsin; trypsin homology  
C:Keywords: hydrolase; serine proteinase; zymogen  
F:1-21/Domain: signal sequence #status predicted <SIG>  
F:22-30/Domain: activation peptide #status predicted <ACT>  
F:31-274/Product: trypsin I #status predicted <MAT>  
F:31-266/Domain: trypsin homology <TRY>  
F:74,120,223/Active site: His, Asp, Ser #status predicted

Query Match 18.3%; Score 101; DB 2; Length 274;  
Best Local Similarity 30.9%; Pred. No. 0.0018;  
Matches 29; Conservative 15; Mismatches 32; Indels 18; Gaps 3;

QY 5 TLKNDQOQIHQP-----SGVVAAPGYMPSGMSANQRBPRTQTSQFDIATVLAQWVN 59  
Db 85 TLKRVSGTHLYQDDQLPVSRIWHPQFY-----IIQT--GADIALLEBPVN 131

QY 60 LQSGIRVSLPQPSDIPPGTGVFIVGYGRDND 93  
Db 132 ISSRVHTVWLPASERFPFGMPCWVTGMDVDND 165

RESULT 7  
A:Residues: 1-275 <VA2>  
A:Cross-references: GB:M33494; NID:G3927804; PIDN:AA083172.1; PID:G339977  
A:Accession: D35863  
A:Molecule type: mRNA  
A:Residues: 1-275 <VA2>  
A:Cross-references: GB:M33491  
A:Butterfield, J.H.; Weller, D.A.; Hunt, L.W.; Wynn, S.R.; Roche, P.C.  
J. Leukoc. Biol. 47, 409-419, 1990  
A:Title: Purification of trypsin from a human mast cell line.  
A:Reference number: A60939; MUID:90244210; PMID:2110591  
A:Accession: A60939  
A:Molecule type: protein  
A:Residues: 31-38, 'P', '40-41', 'X', '43', 'T', '45-48', 'X', '50 <BUT>  
A:Experimental source: mast cell  
A:Note: 44-Gly was also found  
R:Cromlish, J.A.; Seidah, N.G.; Marcinkiewicz, M.; Hamelin, J.; Johnson, D.A.; Chretien, J. Biol. Chem. 262, 1363-1373, 1987  
A:Title: Human pituitary trypsin: molecular forms, NH-2-terminal sequence, immunocytochemistry  
A:Reference number: A39326; MUID:87109258; PMID:3543004  
A:Accession: A39326  
A:Molecule type: protein  
A:Residues: 31-38 <CRO>  
A:Experimental source: pituitary  
C:Genetics:  
A:Introns: 21/1; 78/2; 177/1; 221/3  
C:Superfamily: trypsin; trypsin homology  
C:Keywords: hydrolase; serine proteinase; zymogen  
F:1-21/Domain: signal sequence #status predicted <SIG>  
F:22-30/Domain: activation peptide #status predicted <ACT>  
F:31-275/Product: trypsin I #status predicted <MAT>  
F:31-267/Domain: trypsin homology <TRY>  
F:74,121,224/Active site: His, Asp, Ser #status predicted

Query Match 17.9%; Score 99; DB 2; Length 275;  
Best Local Similarity 39.6%; Pred. No. 0.0031;  
Matches 19; Conservative 11; Mismatches 18; Indels 0; Gaps 0;

QY 46 GFDAIVMLAQWVNLQSGIRVSLPQPSDIPPGTGVFIVGYGRDND 93  
Db 119 GADIALLEBPVNVSSHVHTVTLPPASERFPFGMPCWVTGMDVDND 166

RESULT 8  
A:Residues: 1-275 <VA2>  
A:Cross-references: GB:M33494; NID:G3927804; PIDN:AA083172.1; PID:G339977  
A:Accession: D35863  
A:Molecule type: mRNA  
A:Residues: 1-275 <VA2>  
A:Cross-references: GB:M33493; NID:G339984; PIDN:AAA36780.1; PID:G339985  
C:Superfamily: trypsin; trypsin homology  
C:Keywords: hydrolase; serine proteinase; zymogen  
F:1-21/Domain: signal sequence #status predicted <SIG>  
F:22-30/Domain: activation peptide #status predicted <ACT>  
F:31-275/Product: trypsin I #status predicted <MAT>  
F:74,121,224/Active site: His, Asp, Ser #status predicted

Query Match 17.9%; Score 99; DB 2; Length 275;  
Best Local Similarity 39.6%; Pred. No. 0.0031;  
Matches 19; Conservative 11; Mismatches 18; Indels 0; Gaps 0;

QY 46 GFDAIVMLAQWVNLQSGIRVSLPQPSDIPPGTGVFIVGYGRDND 93  
Db 119 GADIALLEBPVNVSSHVHTVTLPPASERFPFGMPCWVTGMDVDND 166

RESULT 9  
A:Residues: 1-177 <CHU>  
A:Cross-references: EMBL:M68899  
C:Superfamily: trypsin; trypsin homology  
C:Keywords: hydrolase; mast cell; serine proteinase  
F:1-170/Domain: trypsin homology (fragment) <TRY>

Query Match 17.4%; Score 96; DB 2; Length 177;  
Best Local Similarity 39.2%; Pred. No. 0.0039;  
Matches 20; Conservative 8; Mismatches 23; Indels 0; Gaps 0;

QY 48 DIAIVMLAQWVNLQSGIRVSLPQPSDIPPGTGVFIVGYGRDNDPDR 98  
Db 40 DIVLKLKQANLTSADVPLPAPSPAPKPGTWCWAGWGRGTGLKXISR 90

```

RESULT 10
A29934
pancreatic elastase (EC 3.4.21.36) IIIA precursor - human
N:Alternate names: protease E
C:Species: Homo sapiens (man)
C>Date: 15-Dec-1988 #sequence_revision 15-Dec-1988 #text_change 20-Jun-2000
C:Accession: A29934; JX0045
R:Tani, T.; Ohsumi, J.; Mita, K.; Takiguchi, Y.
J. Biol. Chem. 263, 1231-1239, 1988
A>Title: Identification of a novel class of elastase isozyme, human pancreatic elastase
A:Reference number: A92664; MUID:86087253; PMID:2826474
A:Accession: A29934
A:Molecule type: DNA
A:Residues: 1-270 <TAN>
A:Cross-references: GB:J03516
R:Shirai, Y.; Takemura, K.; Yoshida, H.; Sato, Y.; Iijima, H.; Shimada, Y.; Miyayama, T.
J. Biochem. 104, 259-264, 1988
A>Title: Molecular cloning of complementary DNA encoding one of the human pancreatic pro
A:Reference number: JX0045; MUID:89034017; PMID:2460440
A:Accession: JX0045
A:Molecule type: mRNA
A:Residues: 1-240, 'G', 242-270 <SHI>
A:Cross-references: GB:D00306; NID:g220013; PIDN:BA400212.1; PID:g220014
C:Comment: This enzyme is an alanine-specific serine proteinase that has little elastoly
C:Genetics:
A:Introns: 15/1; 43/3; 76/2; 121/2; 167/1; 214/3; 265/3
C:Superfamily: trypsin; trypsin homology
C:Keywords: hydrolase; pancreas; serine proteinase
F:1-15/Domain: signal sequence #status predicted <SIG>
F:16-28/Domain: propeptide #status predicted <PRO>
F:29-270/Product: pancreatic elastase IIIA #status experimental <MAT>
F:73-123,21/Active site: His, Asp, Ser #status predicted

```

```

Query Match 17.3%; Score 95.5; DB 2; Length 270;
Best Local Similarity 29.1%; Pred. No. 0.0073;
Matches 25; Conservative 16; Mismatches 34; Indels 11; Gaps 1;

```

```

QY 4 LTLRNGDGGHMHPSGVAVAPGPMSCMARQRRIAGTGLSPDIAIYMLAQMNLSG 63
DB 90 LAVKSGPEVATYINSEBLVHPLMNRSCVAC-----GNDIALTKLSRAQLGDA 138
QY 64 IRVSLPPSDIPPGTGVFIYGYGR 89
DB 139 VQLASLPAGDILPMKTPCYITGMGR 164

```

```

RESULT 11
A38678
mast cell proteinase-like protein precursor - mouse
C:Species: Mus musculus (house mouse)
C>Date: 16-Sep-1992 #sequence_revision 16-Sep-1992 #text_change 22-Jun-1999
C:Accession: A38678
R:Sevaftin, W.E.; Sullivan, T.P.; Conder, G.A.; Ebrahimi, A.; Marcham, P.; Johnson, S.S.;
J. Biol. Chem. 266, 1934-1941, 1991
A>Title: Cloning of the cDNA and gene for mouse mast cell protease 4. Demonstration of i
es of the mouse and rat.
A:Reference number: A38678; MUID:91107701; PMID:1988455
A:Accession: A38678
A:Molecule type: DNA
A:Residues: 1-246 <SER>
A:Cross-references: GB:M57401; NID:g200514; PIDN:AAA39990.1; PID:g200515
A>Note: the authors translated the codon GGA for residue 15 as Glu
C:Genetics:
A:Introns: 19/1; 69/2; 114/3; 199/3
C:Superfamily: trypsin; trypsin homology
C:Keywords: zymogen
F:1-18/Domain: signal sequence #status predicted <SIG>
F:19-20/Domain: activation peptide #status predicted <ACT>
F:21-246/Product: mast cell proteinase-like protein #status predicted <MAT>
F:21-239/Domain: trypsin homology <TRY>
F:65,109,202/Active site: His, Asp, Ser #status predicted

```

```

Query Match 16.8%; Score 93; DB 2; Length 246;
Best Local Similarity 42.9%; Pred. No. 0.012;
Matches 18; Conservative 7; Mismatches 17; Indels 0; Gaps 0;

```

```

QY 48 DIAIYMLAQMNLSGIRVSLPPGSDIPPGTGVFIYGYGR 89
DB 109 DIVLLKRLKQNLTSADVPLPGPSDPKPGTWCWAGWGR 150

```

```

RESULT 12
B35863
tryptase (EC 3.4.21.59) II precursor - human
N:Alternate names: tryptase beta
C:Species: Homo sapiens (man)
C>Date: 03-Feb-1994 #sequence_revision 03-Feb-1994 #text_change 21-Jul-2000
C:Accession: B35863; A37193; I59473
R:Vanderslice, P.; Ballinger, S.M.; Tam, E.K.; Goldstein, S.M.; Craik, C.S.; Caughey, G.
Proc. Natl. Acad. Sci. U.S.A. 87, 3811-3815, 1990
A>Title: Human mast cell tryptase: multiple cDNAs and genes reveal a multigene serine p
A:Reference number: A35863; MUID:90251647; PMID:2187193
A:Accession: B35863
A:Molecule type: mRNA
A:Residues: 1-275 <VAN>
A:Cross-references: GB:M33492; NID:g339982; PIDN:AAA36779.1; PID:g339983
A>Note: residues 2-275 are derived from mRNA; residue one was inferred from the genomic
R:Miller, J.S.; Moxley, G.; Schwartz, L.B.
J. Clin. Invest. 86, 864-870, 1990
A>Title: Cloning and characterization of a second complementary DNA for human tryptase.
A:Reference number: A37193; MUID:90365005; PMID:2203827
A:Accession: A37193
A:Molecule type: mRNA
A:Residues: 1-275 <MLI>
A:Cross-references: GB:M37488; NID:g179583; PIDN:AAA51843.1; PID:g179584
R:Blom, T.; Hellman, L.
Scand. J. Immunol. 37, 203-208, 1993
A>Title: Characterization of a tryptase mRNA expressed in the human basophil cell line
A:Reference number: I59473; MUID:93166209; PMID:8434231
A:Accession: I59473
A>Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-275 <RBS>
A:Cross-references: GB:S5551; NID:g265666; PIDN:AAD13876.1; PID:g4261576
A:Experimental source: basophil cell line K0812
C:Genetics:
A:Gene: GDB:TPS1
A:Cross-references: GDB:125890; OMIM:191080
A:Map position: 16pter-16qter
C:Superfamily: trypsin; trypsin homology
C:Keywords: hydrolase; serine proteinase; zymogen
F:1-21/Domain: signal sequence #status predicted <SIG>
F:22-30/Domain: activation peptide #status predicted <ACT>
F:31-275/Product: tryptase I #status predicted <MAT>
F:31-267/Domain: trypsin homology <TRY>
F:74,121,224/Active site: His, Asp, Ser #status predicted

```

```

Query Match 16.8%; Score 93; DB 2; Length 275;
Best Local Similarity 37.5%; Pred. No. 0.014;
Matches 18; Conservative 11; Mismatches 19; Indels 0; Gaps 0;

```

```

QY 46 GFDIAIYMLAQMNLSGIRVSLPPGSDIPPGTGVFIYGYGRDND 93
DB 119 GADIALLEBRVKKVSHVHTLPPASBTPPGMPCWVTGWDVND 166

```

```

RESULT 13
B29934
pancreatic elastase (EC 3.4.21.36) IIIB precursor - human
N:Alternate names: 35K glycoprotein; pancreatic protein p35; protein G32; proteinase E
C:Species: Homo sapiens (man)
C>Date: 15-Dec-1988 #sequence_revision 15-Dec-1988 #text_change 22-Jun-1999
C:Accession: B29934; A90516; A3257; A28932; S04999; S04490
R:Tani, T.; Ohsumi, J.; Mita, K.; Takiguchi, Y.

```



J. Biol. Chem. 263, 1231-1239, 1988  
A/Title: Identification of a novel class of elastase isozyme, human pancreatic elastase  
A/Reference number: A92664; MUID:88087253; PMID:2826474  
A/Accession: B29934  
A/Molecule type: mRNA  
A/Residues: 1-270 <TAN>  
A/Cross-references: GB:M18692; NID:9607029; PIDN:AAA58454.1; PID:g182035  
R/Shen, W.; Fletcher, T.S.; Largman, C.  
Biochemistry 26, 3447-3452, 1987  
A/Title: Primary structure of human pancreatic protease E determined by sequence analysis  
A/Reference number: A90516; MUID:88000545; PMID:3477287  
A/Accession: A90516  
A/Molecule type: mRNA  
A/Residues: 'G', 5-63, 'G', 65-78, 'W', 80-118, 'G', 120-163, 'P', 165-270 <SHE>  
R/Fletcher, T.S.  
Submitted to GenBank, August 1987  
A/Reference number: A94507  
A/Contents: revision to residue 119  
A/Accession: A27206  
A/Molecule type: DNA  
A/Residues: 'G', 5-63, 'G', 65-78, 'W', 80-163, 'P', 165-270 <FLB>  
R/Aviles, F.X.; Pascual, R.; Salva, M.; Bonicel, J.; Putgerver, A.  
Biochem. Biophys. Res. Commun. 163, 1191-1196, 1989  
A/Title: Generation of a subunit III-like protein by autolysis of human and porcine pro  
A/Reference number: A33257; MUID:89392022; PMID:2675835  
A/Accession: A33257  
A/Molecule type: Protein  
A/Residues: 18-57 <AVI>  
R/Guy-Crotte, O.; Barthe, C.; Basso, D.; Fournet, B.; Figarella, C.  
Biochem. Biophys. Res. Commun. 156, 318-322, 1988  
A/Title: Characterization of two glycoproteins of human pancreatic juice: P35, a truncat  
A/Reference number: A28932; MUID:89025862; PMID:3178837  
A/Accession: A28932  
A/Molecule type: Protein  
A/Residues: 'X', 32-52, 'X', 54-55, 'XXX', 59-63 <GVY>  
R/Mouland, M.; Kerielec, B.; Mallet, B.; Chapus, C.  
FEBS Lett. 250, 166-170, 1989  
A/Title: Identification of a procarboxypeptidase A-truncated protease E binary complex  
A/Reference number: S04999; MUID:89325560; PMID:2753124  
A/Accession: S04999  
A/Molecule type: Protein  
A/Residues: 31-50 <MOU>  
R/Mendorf, P.; Geyer, R.; Szegoleit, A.; Linder, D.  
FEBS Lett. 249, 275-278, 1989  
A/Title: Localization and characterization of the glycosylation site of human pancreatic  
A/Reference number: S04490; MUID:89289996; PMID:2737288  
A/Accession: S04490  
A/Molecule type: Protein  
A/Residues: 94-128/132-164 <MEN>  
C/Superfamily: trypsin; trypsin homology  
C/Keywords: glycoprotein; hydrolase; serine proteinase; zymogen  
F/1-17/Domain: signal sequence #status predicted <SIG>  
F/18-28/Domain: activation peptide #status predicted <ACT>  
F/29-263/Domain: (or 31-270) pancreatic elastase IIB #status predicted <MAT>  
F/29-263/Domain: trypsin homology <TRY>  
F/73-123/Active site: His, Asp, Ser #status predicted  
F/114/Binding site: carbohydrate (asn) (covalent) #status experimental  
F/153/Binding site: carbohydrate (asn) (covalent) #status absent

Query Match 16.8%; Score 92.5; DB 2; Length 270;  
Best Local Similarity 28.6%; Pred. No. 0.015;  
Matches 24; Conservative 16; Mismatches 33; Indels 11; Gaps 1;

OY 6 LRNDDQGIHHQPSGVKAPGWPSCMSARQRIAGTISGDDIATVLAQWNLQSGIR 65  
DB 92 VKEGEQVIPIINSGLFVHPIIMNRSVCAC-----GNDIALIKLSRGAQLGDAVQ 140

OY 66 VISLPQSPDIPPGTGVIVGYGR 89  
DB 141 LASLPADILPNEIPCTITGWR 164

RESULT 14

A34910  
mast cell proteinase (EC 3.4.21.-) 2 precursor - mouse  
C/Species: Mus musculus (house mouse)  
C/Date: 20-Jul-1990 #sequence\_revision 20-Jul-1990 #text\_change 22-Jun-1999  
C/Accession: A34910  
R/Serafini, W.E.; Reynolds, D.S.; Rogelj, S.; Lane, W.S.; Conder, G.A.; Johnson, S.S.; A  
J. Biol. Chem. 265, 423-429, 1990  
A/Title: Identification and molecular cloning of a novel mouse mucosal mast cell serine  
A/Reference number: A34910; MUID:90094433; PMID:1688433  
A/Accession: A34910  
A/Status: preliminary  
A/Molecule type: mRNA  
A/Residues: 1-244 <SER>  
A/Cross-references: GB:J05177; NID:9200473; PIDN:AAA39972.1; PID:9200474  
C/Superfamily: trypsin; trypsin homology  
C/Keywords: hydrolase; serine proteinase  
F/21-237/Domain: trypsin homology <TRY>  
F/165,109,202/Active site: His, Asp, Ser #status predicted

Query Match 16.2%; Score 89.5; DB 2; Length 244;  
Best Local Similarity 25.4%; Pred. No. 0.029;  
Matches 29; Conservative 15; Mismatches 37; Indels 33; Gaps 3;

OY 5 LRNDDQGIHHQPSGVKAPGWPSCMSARQRIAGTISGDDIATVLAQWNLQSGIR 89  
DB 41 TTKNSKE-----RCGFLIAPQFWMTAHCGRSEISVILGAHNINKNEPTQOIIKTEKTF 96

OY 38 RPIAQTSGF-DIAIWLQWNLQSGIRVLSLPQSPDIPPGTGVIVGYGR 89  
DB 97 VHPKQVLSGRNDIWLKLRKKAELNSDVVISLPSSSDFTKPKMCTAGWK 150

RESULT 15

JC4171  
tryptase (EC 3.4.21.59) precursor - rat  
N/Alternate names: mast cell tryptase  
C/Species: Rattus norvegicus (Norway rat)  
C/Date: 27-Aug-1995 #sequence\_revision 27-Oct-1995 #text\_change 20-Jun-2000  
C/Accession: JC4171  
R/Ide, H.; Itoh, H.; Tomita, M.; Murakumo, Y.; Kobayashi, T.; Maruyama, H.; Osada, Y.;  
U. Biochem. 118, 210-215, 1995  
A/Title: cDNA sequencing and expression of rat mast cell tryptase.  
A/Reference number: JC4171; MUID:96015171; PMID:8537314  
A/Accession: JC4171  
A/Molecule type: mRNA  
A/Residues: 1-274 <IDB>  
A/Cross-references: DDBJ:D38455; NID:9556555; PIDN:BA07486.1; PID:9556556  
C/Comment: This enzyme is basically specific for a connective tissue mast cell, it is u  
elase inhibitors.  
C/Superfamily: trypsin; trypsin homology  
C/Keywords: glycoprotein; hydrolase; mast cell; serine proteinase; zymogen  
F/1-19/Domain: signal sequence #status predicted <SIG>  
F/20-29/Domain: activation peptide #status predicted <ACT>  
F/30-274/Product: mast cell tryptase #status predicted <MAT>  
F/30-266/Domain: trypsin homology <TRY>  
F/73-120/Active site: His, Asp, Ser #status predicted  
F/131/Binding site: carbohydrate (asn) (covalent) #status predicted

Query Match 16.1%; Score 89; DB 2; Length 274;  
Best Local Similarity 41.7%; Pred. No. 0.038;  
Matches 20; Conservative 9; Mismatches 19; Indels 0; Gaps 0;

OY 46 GPDIATVLAQWNLQSGIRVLSLPQSPDIPPGTGVIVGYGRDND 93  
DB 118 GADIALLELEIPNVVSTHPIISLPASPTSPGSCWVTGWDGIDSD 165

RESULT 16

A46721  
chymase (EC 3.4.21.39) precursor - mouse  
N/Alternate names: mast cell proteinase 2; mMCP-2  
C/Species: Mus musculus (house mouse)  
C/Date: 21-Sep-1993 #sequence\_revision 18-Nov-1994 #text\_change 22-Jun-1999

C/Accession: A46721  
R/Guisth, M.F.; Nadeau, J.H.; Johnson, K.R.; McNeil, H.P.; Gratian, K.M.; Austen, K.F.;  
J. Biol. Chem. 268, 11372-11379, 1993  
A/Title: A closely linked complex of mouse mast cell-specific chymase genes on chromosome  
A/Reference number: A46721; MUID:93266596; PMID:8098710  
A/Accession: A46721  
A/Status: preliminary  
A/Molecule type: nucleic acid  
A/Residues: 1-244 <GUR>  
A/Cross-references: GB:L08466; NID:9200967; PIDN:AAA74555.1; PID:9200968  
A/Note: sequence extracted from NCBI backbone (NCBIN:132844, NCBI:P:132845)  
C/Superfamily: trypsin; trypsin homology  
C/Keywords: hydrolase; serine proteinase  
F/21-237/Domain: trypsin homology <TRY>  
F/65,109,202/Active site: His, Asp, Ser #status predicted

Query Match 16.0%; Score 88.5; DB 2; Length 244;  
Best Local Similarity 42.3%; Pred. No. 0.038; Indels 1; Gaps 1;  
Matches 22; Conservative 7; Mismatches 22;

Qy 39 PIAQTLSGF-DIAIVMLAQMVLQSGIRVLSLPPSDIPPGTGVFVGYGR 89  
| | | | | : : : : : | | | | | : : : : :  
Db 99 PKFYLSGFYDMLKLQKKAFLNDVYISLPSSSDFIKPKMCMYAGWK 150

RESULT 17  
113596  
trypsin homolog - fruit fly (Drosophila melanogaster)  
N/Alternate names: protein EG:80H7.1  
C/Species: Drosophila melanogaster  
C/Date: 13-Aug-1999 #sequence\_revision 13-Aug-1999 #text\_change 17-Nov-2000  
A/Accession: T13596  
R/Benos, P.  
submitted to the EMBL Data Library, April 1999  
A/Description: Sequencing the distal X chromosome of Drosophila melanogaster.  
A/Reference number: Z17667  
A/Accession: T13596  
A/Status: preliminary; translated from GB/EMBL/DBJ  
A/Molecule type: DNA  
A/Residues: 1-281 <BEN>  
A/Cross-references: EMBL:AL011027; PIDN:CAA19843.2  
C/Genetics:  
A/Cross-references: FlyBase:FBgn0000481  
A/Introns: 159/3  
A/Note: EG:80H7.1  
C/Superfamily: trypsin; trypsin homology

Query Match 16.0%; Score 88.5; DB 2; Length 281;  
Best Local Similarity 24.1%; Pred. No. 0.045;  
Matches 26; Conservative 17; Mismatches 40; Indels 25; Gaps 3;

Qy 4 LTLRNGDGGIHHOPSGVVA-----PGYMPSCMSARQRPIAQTLSGFDIAIVMLAQM 58  
| | | | | : : : | | | | | : : : | | | | |  
Db 79 LLLQSGQSLAKNNSSGVAVVAALFVHPGEPEDKYV-----DIALQLAQSV 126

Qy 59 NLQSGIRVLSLPPSDIPPGTGVFVGYGRDNDPDRPKNGILKK 106  
| | | | | : : : | | | | | : : : | | | | |  
Db 127 ALSKFVQPRLPDEPRQVTRGNSAVLAAGWGLN-----ATGGVYQQ 166

RESULT 18  
B32410  
mastocytoma proteinase (EC 3.4.21.-) precursor - dog  
C/Species: Canis lupus familiaris (dog)  
C/Date: 12-Oct-1989 #sequence\_revision 31-Dec-1993 #text\_change 10-Sep-1997  
A/Accession: B32410  
R/Vanderlize, P.; Craik, C.S.; Nadel, J.A.; Caughey, G.H.  
Biochemistry 28, 4148-4155, 1989  
A/Title: Molecular cloning of dog mast cell trypsinase and a related protease: structural  
A/Reference number: A32410; MUID:89352460; PMID:2504277  
A/Accession: B32410  
A/Molecule type: mRNA  
A/Residues: 1-269 <VAN>

A/Cross-references: GB:M24665; NID:g163984; PID:g163985; GB:J02862  
A/Note: the authors translated the codon ACG for residue 114 as Ser  
C/Superfamily: trypsin; trypsin homology  
C/Keywords: hydrolase; serine proteinase  
F/1-19/Domain: signal sequence #status predicted <SIG>  
F/20-268/Product: mastocytoma proteinase #status predicted <MAT>  
F/20-259/Domain: trypsin homology <TRY>  
F/66,116,217/Active site: His, Asp, Ser #status predicted

Query Match 15.5%; Score 85.5; DB 2; Length 269;  
Best Local Similarity 38.8%; Pred. No. 0.091;  
Matches 19; Conservative 9; Mismatches 20; Indels 1; Gaps 1;

Qy 48 DIAIVMLAQMVLQSGIRVLSLPPSDIPPGTGVFVGYGRDNDP 96  
| | | | | : : : | | | | | : : : | | | | |  
Db 116 DIALKLKAPLTLSEVDNLVSLPSLIVPPGMLCWYAGWG-DIADHP 163

RESULT 19  
E1KRT1  
pancreatic elastase (EC 3.4.21.36) I precursor - rat  
C/Species: Rattus norvegicus (Norway rat)  
C/Date: 18-Aug-1982 #sequence\_revision 18-Aug-1982 #text\_change 18-Jun-1999  
A/Accession: A00960; A20534  
R/MacDonald, R.J.; Swift, G.H.; Quinto, C.; Swain, W.; Picet, R.L.; Nikovits, W.; Rut  
Biochemistry 21, 1453-1463, 1982  
A/Title: Primary structure of two distinct rat pancreatic preproelastases determined by  
A/Reference number: A00960; MUID:82182967; PMID:6918221  
A/Accession: A00960  
A/Molecule type: mRNA  
A/Residues: 1-266 <MAC>  
A/Cross-references: GB:V01234; NID:g56088; PIDN:CAA24544.1; PID:g56089  
R/Largman, C.  
Biochemistry 22, 3763-3770, 1983  
A/Title: Isolation and characterization of rat pancreatic elastase.  
A/Reference number: A20534; MUID:84000385; PMID:655050  
A/Accession: A20534  
A/Molecule type: protein  
A/Residues: 17-37, 'X', 39-45 <LAR>  
C/Superfamily: trypsin; trypsin homology  
C/Keywords: hydrolase; pancreas; serine proteinase; zymogen  
F/1-16/Domain: signal sequence #status predicted <SIG>  
F/17-26/Domain: activation peptide #status predicted <APT>  
F/27-266/Product: elastase I #status predicted <MP1>  
F/27-259/Domain: trypsin homology <TRY>  
F/71,119,214/Active site: His, Asp, Ser #status predicted

Query Match 15.4%; Score 85; DB 1; Length 266;  
Best Local Similarity 27.6%; Pred. No. 0.1;  
Matches 27; Conservative 19; Mismatches 32; Indels 20; Gaps 3;

Qy 4 LTLRNGDGGIHHOPSGVVA---PGYMPSCMSARQRPIAQTLSGFDIAIVMLAQM 63  
| | | | | : : : | | | | | : : : | | | | |  
Db 88 LQNDGTEGYVSQK--IWHPTMNSNVVA-----GYDIALRLAQSVTLNNY 134

Qy 64 IRVLSLPPSDIPPGTGVFVGYGRDNDPDRPKNG 101  
| | | | | : : : | | | | | : : : | | | | |  
Db 135 VOLAVLPQEGTILANNNPCYITGWGR-----TRTNG 165

RESULT 20  
A47246  
trypsin (EC 3.4.21.59) 2 - mouse  
C/Species: Mus musculus (house mouse)  
C/Date: 21-Sep-1993 #sequence\_revision 18-Nov-1994 #text\_change 21-Jul-2000  
A/Accession: A47246  
R/McNeill, H.P.; Reynolds, D.S.; Schiller, V.; Gildyal, N.; Gurley, D.S.; Austen, K.F.;  
Proc. Natl. Acad. Sci. U.S.A. 89, 11174-11178, 1992  
A/Title: Isolation, characterization, and transcription of the gene encoding mouse mast  
A/Reference number: A47246; MUID:93087489; PMID:1454796  
A/Accession: A47246  
A/Status: preliminary  
A/Molecule type: nucleic acid

A/Residues: 1-273 <MCN>  
A/Cross-References: GB:L00653; NID:G200518; PIDN:AAA39992.1; PID:G200519  
A/Note: Sequence extracted from NCBI backbone (NCBI:119745, NCBI:119746)  
C/Superfamily: trypsin; trypsin homology  
C/Keywords: hydrolase; serine proteinase  
F:29-265/Domain: trypsin homology <TRY>

Query Match 15.0%; Score 83; DB 2; Length 273;  
Best Local Similarity 36.2%; Pred. No. 0.17;  
Matches 17; Conservative 9; Mismatches 21; Indels 0; Gaps 0;

QY 46 GPDIAVWLAQWNLQSGIRVSLPPPSDIPPGTGVIYGYGRDDN 92  
DB 117 GADIALKLITNPVNISDYHVPVLPPEASETPPSGTLCTWGTGNTDN 163

## RESULT 21

PRRTG  
mast cell proteinase II (EC 3.4.21.-) precursor - rat  
N/Alternate names: group-specific proteinase

C/Species: Rattus norvegicus (Norway rat)  
C/Date: 30-Nov-1980 #sequence\_revision 19-Oct-1995 #text\_change 18-Jun-1999

C/Accession: A29548; A00957  
R/Bentley, P.N.; Yin, F.H.; Leder, P.  
J. Biol. Chem. 262, 5377-5384, 1987

A/Title: Cloning of the mast cell protease, RMCP II. Evidence for cell-specific expression  
A/Reference number: A92636; MUID:87165980; PMID:3549719

A/Accession: A29548  
A/Molecule type: DNA; mRNA

A/Residues: 1-247 <BEN>

A/Cross-References: GB:J02712; NID:G205335; PIDN:AAA6284.1; PID:G205336  
R/Woodbury, R.G.; Katunuma, N.; Kobayashi, K.; Titani, K.; Neurath, H.  
Biochemistry 17, 811-819, 1978

A/Title: Covalent structure of a group-specific proteinase from rat small intestine.  
A/Reference number: A00957; MUID:76124137; PMID:629933

A/Accession: A00957  
A/Molecule type: protein

A/Residues: 21-237, 'T', 239-244 <WOO>  
A/Experimental source: small intestine

C/Comment: This enzyme has chymotrypsin-like specificity toward small substrates. It is the gut and lung.  
C/Genetics:

A/Intons: 19/1; 69/2; 124/3; 199/3  
C/Superfamily: trypsin; trypsin homology

C/Keywords: hydrolase; mast cell; serine proteinase  
F:1-20/Domain: signal sequence #status predicted <SIG>

F:21-247/Product: mast cell proteinase II #status predicted <MAT>  
F:21-239/Domain: trypsin homology <TRY>

F:50-66;143-208;174-187/Disulfide bonds: #status predicted  
F:65;109;202/Active site: His, Asp, Ser #status predicted

Query Match 14.8%; Score 81.5; DB 1; Length 247;  
Best Local Similarity 36.0%; Pred. No. 0.23;  
Matches 18; Conservative 9; Mismatches 22; Indels 1; Gaps 1;

QY 48 DIATVLAQWNLQSGIRVSLPPPSDIPPGTGVIYGYGRDDN 97  
DB 109 DIMLKKEKVELTPAVNVPLPSPDFIRGAMCAAGMK-TEVRDPT 157

## RESULT 22

JE0315  
low-density lipoprotein receptor-related protein - mouse  
C/Species: Mus musculus (house mouse)

C/Date: 16-Jul-1999 #sequence\_revision 16-Jul-1999 #text\_change 18-Aug-2000  
C/Accession: JE0315

R/Tomita, Y.; Kim, D.; Magoori, K.; Fujino, T.; Yamamoto, T.T.  
J. Biochem. 124, 784-789, 1998

A/Title: A novel low-density lipoprotein receptor protein with type II membrane  
A/Reference number: JE0315; MUID:96425596; PMID:9756624

A/Accession: JE0315  
A/Status: preliminary  
A/Molecule type: mRNA

A/Residues: 1-1113 <TCM>  
A/Cross-References: DDBJ:AB013874; NID:G3869144; PIDN:BA434371.1; PID:G3869145  
C/Superfamily: trypsin homology; LDL receptor ligand-binding repeat homology  
F:337-372/Domain: LDL receptor ligand-binding repeat homology <LDL1>  
F:374-408/Domain: LDL receptor ligand-binding repeat homology <LDL2>  
F:410-445/Domain: LDL receptor ligand-binding repeat homology <LDL3>  
F:447-482/Domain: LDL receptor ligand-binding repeat homology <LDL4>  
F:468-682/Domain: LDL receptor ligand-binding repeat homology <LDL5>  
F:684-720/Domain: LDL receptor ligand-binding repeat homology <LDL6>  
F:723-757/Domain: LDL receptor ligand-binding repeat homology <LDL7>  
F:863-1097/Domain: trypsin homology <TRY>

Query Match 14.7%; Score 81; DB 2; Length 1113;  
Best Local Similarity 28.0%; Pred. No. 1.5;  
Matches 23; Conservative 14; Mismatches 39; Indels 6; Gaps 2;

QY 13 GIHH-QPSGVVAPGVVPSGMSAQRPIAQTLSGPDIAVWLAQWNLQSGIRVSLP 70  
DB 926 GINNLDHPSGFVQTRFVKTILLPRYSRAVVD---YDISVELSDINETSYPVCLP 981

QY 71 QPSDIPPGTGVIYGYGRDDN 92  
DB 982 SPEYLBPDTCTYTGGMGMGN 1003

## RESULT 23

J01473  
pancreatic elastase (EC 3.4.21.36) IV precursor - rat  
C/Species: Rattus norvegicus (Norway rat)

C/Date: 17-Jul-1992 #sequence\_revision 17-Jul-1992 #text\_change 22-Jun-1999  
C/Accession: J01473; S23783

R/Kang, J.; Wiegand, U.; Mueller-Hill, B.  
Gene 110, 181-187, 1992

A/Title: Identification of cDNAs encoding two novel rat pancreatic serine proteases.  
A/Reference number: J01471; MUID:92165057; PMID:1537555

A/Accession: J01473  
A/Molecule type: mRNA

A/Residues: 1-268 <KAN>  
A/Cross-References: EMBL:X59014; NID:G56090; PIDN:CAA41753.1; PID:G56091

C/Superfamily: trypsin; trypsin homology  
C/Keywords: hydrolase; serine proteinase; zymogen

F:1-16/Domain: signal sequence #status predicted <SIG>  
F:17-29/Domain: activation peptide #status predicted <ACT>

F:30-268/Product: pancreatic elastase IV #status predicted <MAT>  
F:30-262/Domain: trypsin homology <TRY>

F:74;121;216/Active site: His, Asp, Ser #status predicted

Query Match 14.5%; Score 80; DB 2; Length 268;  
Best Local Similarity 36.4%; Pred. No. 0.36;  
Matches 16; Conservative 10; Mismatches 18; Indels 0; Gaps 0;

QY 46 GPDIAVWLAQWNLQSGIRVSLPPPSDIPPGTGVIYGYGR 89  
DB 119 GTDIALIKLAEVLSNITQVACIPREGSLPDPYCVYTGWGR 162

## RESULT 24

A47136  
macrophage-stimulating protein 1 precursor - human  
C/Species: Homo sapiens (man)

C/Date: 03-May-1994 #sequence\_revision 14-Nov-1997 #text\_change 18-Jun-1999  
C/Accession: A40331; A47136; A61395

R/Han, S.; Stuart, L.A.; Degen, S.J.F.  
Biochemistry 30, 9768-9780, 1991

A/Title: Characterization of the DNFI552 locus on human chromosome 3: identification of  
A/Reference number: A40331; MUID:92002016; PMID:1655021

A/Accession: A40331  
A/Molecule type: DNA

A/Residues: 1-711 <HA1>  
A/Cross-References: GB:M74179  
A/Accession: B40331  
A/Molecule type: mRNA  
A/Residues: 1-711 <HA2>

A:Cross-references: GB:M74178; NID:G183976; PIDN:AAA50165.1; PID:G183977  
R.Yoshimura, T.; Yuhki, N.; Wang, M.H.; Skeeel, A.; Leonard, E.J.  
J. Biol. Chem. 268, 15461-15468, 1993  
A:Title: Cloning, sequencing, and expression of human macrophage stimulating protein (MS  
A:Reference number: A67136; MUID:93340141; PMID:8393443  
A:Accession: A67136  
A:Molecule type: mRNA  
A:Residues: 1-12, 'C', '14-622', 'F', '624-711 <YOS>  
A:Cross-references: GB:LI1924; NID:G398037; PIDN:AAA59672.1; PID:G398038  
A:Note: authors translated the codon TTT for residue 623 as Ieu; parts of this sequence  
R.Skeel, A.; Yoshimura, T.; Showalter, S.D.; Tanaka, S.; Appella, E.; Leonard, E.J.  
J. Exp. Med. 173, 1227-1234, 1991  
A:Title: Macrophage stimulating protein: purification, partial amino acid sequence, and  
A:Reference number: A61395; MUID:91217635; PMID:18271141  
A:Accession: A61395  
A:Molecule type: protein  
A:Residues: 230-247, 286-291, 'E', '293-295', 'X', '297-301', 'X', '303', 'E', '305', 'EX', '308-310, 326-331  
A:Experimental source: plasma  
A:Genetic8:  
A:Gene: GDB:MST1; D3P1582; DNFI582; HGFL  
A:Cross-references: GDB:128833; OMIM:142408  
A:Map position: 3p21-3p21.3  
C:Complex: disulfide-bonded heterodimer of chains derived from the same precursor  
C:Superfamily: hepatocyte growth factor; kringle homology; trypsin homology  
C:Keywords: duplication; glycoprotein; growth factor; kringle; plasma  
F:1-18/Domain: signal sequence #status predicted <SIG>  
F:19-483, 484-711/Product: macrophage-stimulating protein 1 #status predicted <MAT>  
F:19-483/Domain: alpha chain #status predicted <ACH>  
F:119-186/Domain: kringle homology <KR1>  
F:191-268/Domain: kringle homology <KR2>  
F:283-361/Domain: kringle homology <KR3>  
F:370-448/Domain: kringle homology <KR4>  
F:484-711/Domain: beta chain #status predicted <BCH>  
F:484-704/Domain: trypsin homology <TRY>  
F:56-78, 60-66, 110-186, 131-161, 157-181, 191-268, 212-251, 240-263, 283-361, 304-343, 332-355, 37  
F:72, 296, 615/Binding site: carbohydrate (asn) (covalent) #status predicted

Query Match 14.5%; Score 80; DB 1; Length 711;  
Best Local Similarity 26.9%; Pred. No. 1.1;  
Matches 32; Conservative 17; Mismatches 30; Indels 40; Gaps 6;

OY 4 LTRNGDQGGH-----HOP-SGVAVAPGYM-----PSGM 32  
DB 496 VSLRN--RQGHFCGSLVKEQWILTARQCFSSCHMPLTGYVWLTGLPQNPGHSPSL- 552  
OY 33 SARORRPIAQLT---SGFDIAIWMALQWNLQSGIRVLSLPQPSDIPPGTGVTIVGYG 88  
DB 553 ---QRPVAKWVCGPSGSQVLKLRSVTLNQRVALICLPPEWYVVPPTKCEIAGWG 608

RESULT 25  
T30337  
polyprotein - African clawed frog  
C:Species: Xenopus laevis (African clawed frog)  
C:Date: 22-Oct-1999 #sequence\_revision 22-Oct-1999 #text\_change 22-Oct-1999  
C:Accession: T30337  
R.Yang, J.C.; Lindsey, L.L.; Hedrick, J.L.  
submitted to the EMBL Data Library, March 1998  
A:Description: cDNA cloning of ovochymase, a chymotrypsin-like protease released from Xe  
A:Reference number: Z20829  
A:Accession: T30337  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: mRNA  
A:Residues: 1-1524 <YAN>  
A:Cross-references: EMBL:U81290; NID:G2981640; PID:G2981641; PIDN:AA24717.1

Query Match 14.5%; Score 80; DB 2; Length 1524;  
Best Local Similarity 28.0%; Pred. No. 2.7;  
Matches 14; Conservative 14; Mismatches 22; Indels 0; Gaps 0;

OY 48 DIAIWMALQWNLQSGIRVLSLPQPSDIPPGTGVTIVGYGDDNDRDPDS 97  
DB 672 DIAIWMALQWNLQSGIRVLSLPQPSDIPPGTGVTIVGYGDDNDRDPDS 721

RESULT 26  
A28566  
T-cell suppressor factor, phosphorylcholine-specific - mouse (fragment)  
C:Species: Mus musculus (house mouse)  
C:Date: 15-Dec-1988 #sequence\_revision 30-Sep-1991 #text\_change 22-Jun-1999  
C:Accession: A28566  
R.Yamasaki, N.; Sugimura, K.; Hida, M.; Naito, T.; Watanabe, T.  
Eur. J. Immunol. 17, 247-253, 1987  
A:Title: Sequence analysis of a cDNA clone of a gene encoding a component of a putative  
A:Reference number: A28566; MUID:87162073; PMID:3493908  
A:Accession: A28566  
A:Molecule type: mRNA  
A:Residues: 1-236 <YAM>  
A:Cross-references: GB:M7347; NID:G200267; PIDN:AAA39901.1; PID:G200268  
A:Experimental source: T-cell hybridoma  
A:Note: the authors translated the codon AAC for residue 69 as Val, GTG for residue 74  
as His, and CTC for residue 222 as Val  
C:Superfamily: trypsin; trypsin homology  
F:1-229/Domain: trypsin homology (fragment) <TRY>

Query Match 14.2%; Score 78.5; DB 2; Length 236;  
Best Local Similarity 32.8%; Pred. No. 0.46;  
Matches 19; Conservative 14; Mismatches 18; Indels 7; Gaps 1;

OY 44 LSGFDIAIWMALQWNLQSGIRVLSLPQPSDIPPGTGVTIVGYGDDNDRDPDS 101  
DB 85 VAGDIALVRLAKSVTLNNYVQLGVLPREGTILANNSPCYITGMGR-----TRTNG 135

RESULT 27  
S70439  
pancreatic elastase I (allele HELI-16) probable splice form I - human  
C:Species: Homo sapiens (man)  
C:Date: 03-Aug-1998 #sequence\_revision 03-Aug-1998 #text\_change 16-Dec-1998  
C:Accession: S70439  
R.Kawashima, I.; Tanai, T.; Mita-Honjo, K.; Shimoda-Takano, K.; Ohmine, T.; Furukawa, H.  
DNA Seq. 2, 303-312, 1992  
A:Title: Genomic organization of the human homologue of the rat pancreatic elastase I g  
A:Reference number: A56615; MUID:92338395; PMID:1633328  
A:Accession: S70439  
A:Status: preliminary; not compared with conceptual translation  
A:Molecule type: DNA  
A:Residues: 1-258 <KAM>  
C:Superfamily: trypsin; trypsin homology  
F:19-251/Domain: trypsin homology <TRY>

Query Match 14.2%; Score 78.5; DB 2; Length 258;  
Best Local Similarity 33.3%; Pred. No. 0.51;  
Matches 19; Conservative 13; Mismatches 18; Indels 7; Gaps 1;

OY 45 SGFDIAIWMALQWNLQSGIRVLSLPQPSDIPPGTGVTIVGYGDDNDRDPDS 101  
DB 108 AGYDIALRLAQSIVLNSYVQLGVLPQEGAILANNSPCYITGMGR-----TKTNG 157

RESULT 28  
A56615  
probable pancreatic elastase (EC 3.4.21.36) pseudogene - human  
N:Alternate names: pancreatic elastase I homolog; pancreatic elastase I allele HELI-16,  
C:Species: Homo sapiens (man)  
C:Date: 11-Aug-1995 #sequence\_revision 17-Aug-1995 #text\_change 14-Aug-1998  
C:Accession: A56615; S70440  
R.Kawashima, I.; Tanai, T.; Mita-Honjo, K.; Shimoda-Takano, K.; Ohmine, T.; Furukawa, H.  
DNA Seq. 2, 303-312, 1992  
A:Title: Genomic organization of the human homologue of the rat pancreatic elastase I g  
A:Reference number: A56615; MUID:92338395; PMID:1633328  
A:Accession: A56615  
A:Molecule type: DNA  
A:Residues: 1-267 <KAM>  
A:Cross-references: EMBL:X62259; NID:G31246; EMBL:X62258; GB:S40923; NID:G31247; EMBL:X  
5; GB:S40856; NID:G31251; EMBL:X62256; GB:S40857; NID:G31252; EMBL:X62257; GB:S40859; N

F;31-264/Domain: trypsin homology <TRY>  
F;75,123,218/Active site: His, Asp, Ser #status predicted

A1:Molecule type: DNA  
A1:Residues: 1-257 <HER>

R:Gershenfeld H.K.; Weissman, I.L.  
 Science 232, 854-858, 1986  
 A:Title: Cloning of a cDNA for a T cell-specific serine protease from a cytotoxic T lymphocyte  
 A:Reference numbers: A47550, MUID:66208119, PMID:2422755  
 A:Accession: A47550  
 A:Status: preliminary  
 A:Molecule type: mRNA  
 A:Residues: 12-257 <GER>  
 A:Cross-references: GB:M13226  
 C:Superfamily: trypsin; trypsin homology  
 C:Keywords: hydrolase; serine protease  
 F:26-249/Domain: trypsin homology <TRY>

Query Match 13.6% Score 75; DB 2; Length 257;  
 Best Local Similarity 30.0%; Pred. No. 1,2; Indels 0; Gaps 0;  
 Matches 15; Conservative 10; Mismatches 25;

Qy 48 DIATMLAQMWNLOSGRIVSLPQPSDIPPGTGVFIYGYGDDNDRDS 97  
 Db 110 DLQVRLKKKATVNNRNVALHLPKKGDVVKPQTRCRVAVGWGFGKSAVS 159

### RESULT 33

Granzyme A (E03.4.21.78) precursor, splice form 1 - mouse  
 N;Alternate names: cytotoxic T-cell-specific proteinase I  
 C;Species: Mus musculus (house mouse)  
 C;Date: 30-Sep-1993 #sequence revision 30-Sep-1993 #text change 19-May-2000  
 A;Accession: A45061; S26184; S10085; A27640; A26944; I49703  
 J;Herzberg, R.J.; Gerstenfeld, H.K.; Weissman, I.L.; Su, L.  
 J. Biol. Chem. 267, 25488-25493, 1992  
 A;Title: Genomic organization of the mouse granzyme A gene. Two mRNAs encode the same ma  
 A;Reference number: A45061; MUID:93094270; PMID:1460043  
 A;Accession: A45061  
 A;Status: preliminary  
 A;Molecule type: DNA  
 A;Residues: 1-260 <HR>  
 A;Cross-references: GB:L01429; GB:L00631; NID:g1321600; PIDN:AAA9898.1; PID:g1321603  
 R;Ebner, K.; Kramer, M.D.; Simon, M.M.  
 Genomics 13, 502-508, 1992  
 A;Title: Organization of the gene encoding the mouse T-cell-specific serine proteinase 'A'  
 A;Reference number: S26184; MUID:92347841; PMID:1639378  
 A;Accession: S26184  
 A;Molecule type: DNA  
 A;Residues: 1-260 <EBN>  
 A;Cross-references: EMBL:X60310  
 R;Bogenberger, J.; Haas, M.  
 Oncogene Res. 3, 301-312, 1988  
 A;Title: cDNA clones from autocrine thymic lymphoma cells encode two mitogenic proteins  
 A;Reference number: S10085; MUID:89144553; PMID:2976140  
 A;Accession: S10085  
 A;Molecule type: mRNA  
 A;Residues: 1-260 <BOG>  
 A;Cross-references: EMBL:X14799; NID:G54159; PIDN:CAA32905.1; PID:G54160  
 R;Simon, H.G.; Frutsh, U.; Eckerskorn, C.; Lotzspeich, F.; Kramer, M.D.; Nerz, G.; Simon,  
 Eur. J. Immunol. 18, 855-861, 1988  
 A;Title: Induction of T cell serine proteinase 1 (TSP-1)-specific mRNA in mouse T lymphoc  
 A;Reference number: A27640; MUID:88255076; PMID:3260181  
 A;Accession: A27640  
 A;Molecule type: protein  
 A;Residues: 29-46 <STM>  
 R;Maasson, D.; Tschopp, J.  
 Cell 49, 679-685, 1987  
 A;Title: A family of serine esterases in lytic granules of cytolytic T lymphocytes.  
 A;Reference number: A50894; MUID:87215932; PMID:555842  
 A;Accession: A26944  
 A;Molecule type: protein  
 A;Residues: 29-48 <MAS>  
 R;Jenne, D.E.; Tschopp, J.  
 Immunol. Rev. 103, 53-71, 1988  
 A;Title: Granzymes, a family of serine proteases released from granules of cytolytic T l  
 A;Reference number: I49703; MUID:88272336; PMID:3292936  
 A;Accession: I49703

A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: mRNA  
A:Residues: 1-37 <RES>  
A:Cross-references: GB:M26183; NID:g193623; PION:AAA37735.1; PID:g193624  
C:Comment: Cytoplasmic granules of cytolytic T-lymphocytes contain a family of highly h  
C:Superfamily: trypsin; trypsin homology  
C:Keywords: hydrolase; serine proteinase, T-cell  
F:29-252/Domain: trypsin homology <TRY>  
F:69,113,211/Active site: His, Asp, Ser #status predicted

Query Match	13.6%	Score 75;	DB 2;	Length 260;
Best Local Similarity	30.0%	Pred. No. 1.2		
Matches	15;	Conservative	10;	Mismatches 25; Indels 0; Gaps 0;

  

Qy	48	DIAIYMLAQMNVNLSGIRVYISLPQSDIIPPTGTVYVYGYGRRDDNDRDPDS	97
db	113	DLOVRLKPKKATVNNRNVAIIILPKKGGDVCKKGTGRVAVGWGRFNGKSAAPS	162

### RESULT 34

pancreatic elastase (EC 3.4.21.36) isoform 2 precursor - human  
N,Alternate names: caldesin isoform 2  
C,Species: Homo sapiens (man)  
C,Cdate: 04-Dec-1997 #sequence\_revision 12-Dec-1997 #extc change 30-Jan-1998  
C,Accession: S68826  
P,Totomura, A.; Akiyama, M.; Itoh, H.; Yoshino, I.; Totomura, M.; Nishii, Y.; Noikura, F.BS Lett. 386, 26-28, 1996  
A,Title: Molecular cloning and expression of human caldesin.  
A,Reference number: S68825, MUID:96221265, PMID:8635596  
A,Accession: S68826  
A,Molecule type: mRNA  
A,Residues: 1-268 <TOM>  
A,Experimental source: pancreas  
A,Superfamily: trypsin; trypsin homology  
C,Keywords: hydrolase; pancreas; serine proteinase; zymogen  
P,1-16/Domain: signal sequence #status predicted <Sig>  
P,17-29/Domain: propeptide #status predicted <PRO>  
P,30-266/Product: pancreatic elastase isoform 2 #status predicted <MAT>  
P,30-262/Domain: trypsin homology <TRY>

Query Match	13.6%	Score 75	DB 2	Length 268
Best Local Similarly	35.7%	Pred. No. 1.3		
Matches 15; Conservative	9	Mismatches 18	Indels 0	Gaps 0

Qy 48 DIAIVMLAQMVNLOSGRVLSLPQSPDIPPPGTGVFIVGYGR 89  
|||::|||:-||::|  
Db 121 DIAIKLAHEVELSDPTIQVACLPEKDSLLPKDYPCYVTGWGR 1622  
|||::|||:-||::|

### RESULT 35

pancreatic elastase (EC 3.4.21.36) isoform 1 precursor - human  
N|Alternative names: caldesirin isoform 1  
C|Species: Homo sapiens (man)  
C|Date: 04-Dec-1997 #sequence\_revision 12-Dec-1997 #ext\_change 28-May-1999  
C|Accession: S68825  
R|Tomomura, A.; Akiyama, M.; Itoh, H.; Yoshino, I.; Tomomura, M.; Nishi, Y.; Noikura  
PEBS Lett. 386, 26-28, 1996  
A|Title: Molecular cloning and expression of human caldesirin.  
A|Reference number: S68825; MUID:96221265; PMID:8635596  
A|Accession: S68825  
A|Molecule type: mRNA  
A|Residues: 1-268 <TOM>  
A|Cross-references: GB:S682198; NID:g1839466; P|IDN:AA87104.1; P|ID:g1839467  
A|Experimental source: pancreas  
C|Superfamily: trypsin; trypsin homology  
C|Keywords: hydrolase; pancreas; serine proteinase; zymogen  
F.1-16/Domain: signal sequence #status predicted <SIG>  
F.17-29/Domain: propeptide #status predicted <PRO>  
F.30-268/Product: pancreatic elastase isoform 1 #status predicted <MAT>  
F.30-262/Domain: trypsin homology <TRY>

Query Match 13.6%; Score 75; DB 2; Length 266;  
Best Local Similarity 35.7%; Pred. No. 1.3;  
Matches 15; Conservative 9; Mismatches 18; Indels 0; Gaps 0;  
Cy 48 DIAIVMLAQMNLQSGIRVLSLPSPSDIPPGTGVPIYGVGR 89  
Db 121 DIALIKLAHVELSDPTQVACLPKXDSLLPKDYPCYVTGWGR 162

RESULT 36  
A46504  
Chymase (EC 3.4.21.39) precursor - mouse  
N:Alternate names: mast cell proteinase I  
C:Species: Mus musculus (house mouse)  
C>Date: 18-Jun-1993 #sequence revision 27-Jun-1994 #text\_change 18-Jun-1999  
R:Accession: A46504; S26041; A30105; S35607  
R:Chitidial, N.; McNeill, H.P.; Seebach, S.; Austen, K.F.; Silberstein, D.; Gurish, M.  
J. Immunol. 149, 2123-2129, 1992  
A:Title: IL-10 induces transcription of the gene for mouse mast cell protease-1, a serin  
A:Reference number: A46504; MUID:92388686; PMID:1517575  
A:Accession: A46504  
A:Molecule type: DNA  
A:Residues: 1-246 <GHI>  
A:Cross-references: GB:S44609; NID:9255141; PIDN:AA523194.1; PID:9255142  
A:Experimental source: rIL-10 treated bone marrow  
A:Note: Sequence extracted from NCBI backbone (NCBI:113144)  
R:Hang, R.; Blom, T.; Hellman, L.  
Eur. J. Immunol. 21, 1611-1621, 1991  
A:Title: Cloning and structural analysis of MMCP-1, MMCP-4 and MMCP-5, three mouse mast  
A:Reference number: S26041; MUID:91285010; PMID:2060576  
A:Accession: S26041  
A:Molecule type: DNA  
A:Residues: 1-246 <HTA>  
A:Cross-references: EMBL:X68803; NID:953154; PIDN:CAA48703.1; PID:953155  
A:Experimental source: Strain Balb/c  
R:Tiong, H.L.; Newlands, G.F.J.; Miller, H.R.P.; Charbonneau, H.; Neurath, H.; Woodbury,  
Biochemistry 28, 391-395, 1989  
A:Title: Amino acid sequence of a mouse mucosal mast cell protease.  
A:Reference number: A30105; MUID:89207558; PMID:2706264  
A:Accession: A30105  
A:Molecule type: Protein  
A:Residues: 21-246 <TRO>  
R:Newlands, G.F.J.; Knox, D.P.; Pirie-Shepherd, S.R.; Miller, H.R.P.  
Biochem. J. 294, 127-135, 1993  
A:Title: Biochemical and immunological characterization of multiple glycoforms of mouse  
A:Reference number: S35607; MUID:93371351; PMID:8363563  
A:Accession: S35607  
A:Status: preliminary  
A:Molecule type: protein  
A:Residues: 21-49 <NEW>  
C:Genetics:  
A:Gene: MMCP-1  
A:Introns: 20/1; 70/2; 115/3; 200/3  
C:Superfamily: trypsin; trypsin homology  
C:Keywords: glycoprotein; hydrolase; mast cell; serine proteinase  
F:1-18/Domain: signal sequence #status predicted <SIG>  
F:19-20/Domain: propeptide #status predicted <PRO>  
F:21-246/Product: chymase #status experimental <MAT>  
F:21-233/Domain: trypsin homology <TRY>  
F:50-66,143-208,174-187/Disulfide bonds: #status predicted  
F:65,109,202/Active site: His, Asp, Ser #status predicted  
F:102/Binding site: carbohydrate (asn) (covalent) #status predicted

Query Match 13.5%; Score 74.5; DB 1; Length 246;  
Best Local Similarity 31.4%; Pred. No. 1.3;  
Matches 16; Conservative 12; Mismatches 22; Indels 1; Gaps 1;  
Cy 47 FDIAIVMLAQMNLQSGIRVLSLPSPSDIPPGTGVPIYGVGRDNDHRPS 97  
Db 108 YDMLKLEKRAELTPYDVIFLPSPSDFIDGKMCWTAGWK-TGEKEPT 157

RESULT 37

S64707  
Chymase (EC 3.4.21.39) 1 precursor - Mongolian jird  
N:Alternate names: mast cell protease-1 precursor  
C:Species: Meriones unguiculatus (Mongolian jird)  
C>Date: 28-Oct-1996 #sequence\_revision 13-Mar-1997 #text\_change 20-Jun-2000  
R:Accession: S64707  
R:Itch, H.; Murakumo, Y.; Tomita, M.; Ide, H.; Kobayashi, T.; Maruyama, H.; Horii, Y.;  
Biochem. J. 314, 923-929, 1996  
A:Title: Cloning of the cDNAs for mast-cell chymases from the jejunum of Mongolian gerb  
t cells of mice and rats.  
A:Reference number: S64707; MUID:96177868; PMID:8615790  
A:Accession: S64707  
A:Status: preliminary  
A:Molecule type: mRNA  
A:Residues: 1-246 <TRO>  
A:Cross-references: EMBL:D45173; NID:9633077; PIDN:BA08121.1; PID:9633078  
C:Superfamily: trypsin; trypsin homology  
C:Keywords: hydrolase; serine proteinase  
F:1-8/Domain: signal sequence #status predicted <SIG>  
F:19-20/Domain: propeptide #status predicted <PRO>  
F:21-246/Product: chymase 1 #status predicted <MAT>  
F:21-239/Domain: trypsin homology <TRY>  
F:65,109,202/Active site: His, Asp, Ser #status predicted

Query Match 13.5%; Score 74.5; DB 2; Length 246;  
Best Local Similarity 28.8%; Pred. No. 1.3;  
Matches 19; Conservative 13; Mismatches 25; Indels 9; Gaps 1;  
Cy 33 SARQRRPAAQLTSG-----FDIAIVMLAQMNLQSGIRVLSLPSPSDIPPGTGVF 83  
Db 85 STQOKIRVAKQIAHPSYSPYSLNLDMLKLQKAKVAVSDVLSLPSPSDFINQKVC 144

84 IVYGR 89  
145 AAGWR 150

RESULT 38  
ELPG  
pancreatic elastase (EC 3.4.21.36) 1 precursor - pig  
C:Species: Sus scrofa domestica (domestic pig)  
C>Date: 24-Apr-1984 #sequence\_revision 30-Sep-1990 #text\_change 16-Jun-2000  
R:Accession: J50013; A26777; A100959  
R:Shirau, Y.; Yoshida, H.; Mikayama, T.; Matsuki, S.; Tanaka, J.I.; Ikenaga, H.  
J. Biochem. 99, 1707-1712, 1986  
A:Title: Isolation and expression in Escherichia coli of a cDNA clone encoding porcine  
A:Reference number: A92005; MUID:86304235; PMID:3528137  
A:Accession: J50013  
A:Molecule type: mRNA  
A:Residues: 1-266 <SHI>  
A:Cross-references: GB:X04036; GB:D00070; GB:N00070; NID:91941; PIDN:CAA27670.1; PID:91  
R:Tani, T.; Kawashima, I.; Furukawa, H.; Ohmine, T.; Takiguchi, Y.  
J. Biochem. 101, 591-599, 1987  
A:Title: Characterization of a silent gene for human pancreatic elastase I: structure o  
A:Reference number: A26777; MUID:87250343; PMID:3648024  
A:Accession: A26777  
A:Molecule type: mRNA  
A:Residues: 1-125; 'G', 127-183, 'U', 185-266 <TAN>  
A:Cross-references: GB:D00160; NID:9217683; PIDN:BA00118.1; PID:9217684  
A:Note: the authors translated the codon GGC for residue 58 as Gln, GGC for residue 126  
R:Shotton, D.M.; Hartley, B.S.  
Biochem. J. 131, 643-675, 1973  
A:Title: Evidence for the amino acid sequence of porcine pancreatic elastase.  
A:Reference number: A90267; MUID:73229121; PMID:4578945  
A:Accession: A10061  
A:Molecule type: protein  
A:Residues: 27-91, 'N', 93-203, 'N', 205-266 <SHO>  
R:Shotton, D.M.; Hartley, B.S.  
Nature 225, 811-816, 1970  
A:Title: Three-dimensional structure of tosyl-elastase.  
A:Reference number: A93160; MUID:70114044; PMID:5415110  
A:Contents: annotation; X-ray crystallography, 3.5 angstroms; active site  
C:Superfamily: trypsin; trypsin homology





GenCore version 5.1.4\_p5\_4578  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: April 8, 2003, 10:21:07 ; Search time 2211.23 Seconds  
(without alignments)  
1395.105 Million cell updates/sec

Title: US-10-020-441-2

Perfect score: 552  
Sequence: 1 VSFLRLNGDQGIHHQPSG.....YGRDDNDPRSRKNGIILKK 106

Scoring table: BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 2054640 seqs, 14551402878 residues  
Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Command line parameters:  
-MODL=frame+ p2n.model -DEV=xlh  
-Q=/cgn2.1/USPTO/US10020441/runat\_02042003\_093012\_20970/app\_query.fasta\_1.462  
-DB=genemb1 -QFMT=fastap -SUFFIX=rge -MINMATCH=0.1 -LOOFCU=0 -LOOFEVT=0  
-UNITS=bits -START=1 -END=-1 -MATRIX=blomsum62 -TRANS=human40.cdi -LIST=45  
-DOCALLIGN=200 -THR SCORE=PCT -THR MAX=100 -THR MIN=0 -ALIGN=40 -MODE=LOCAL  
-OUTFMT=ptc -NORM=ext -HEADSIZE=500 -MINLEN=0 -MAXLEN=200000000  
-USER=US10020441 @CGN 1.1 1758 @runat\_02042003\_093012\_20970 -NCFU=6 -ICPU=3  
-NO\_XLPRY -NO\_MMAR -LARGEBUFFER -NEG\_SCORES=0 -WAIT -LONGLOG -DEV\_TIMEOUT=120  
-WARN\_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7  
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : GenEmb1:  
1: gb\_ba:\*  
2: gb\_hcg:\*  
3: gb\_in:\*  
4: gb\_om:\*  
5: gb\_ov:\*  
6: gb\_pat:\*  
7: gb\_ph:\*  
8: gb\_pl:\*  
9: gb\_pr:\*  
10: gb\_ro:\*  
11: gb\_srs:\*  
12: gb\_sy:\*  
13: gb\_un:\*  
14: gb\_vl:\*  
15: em\_da:\*  
16: em\_fun:\*  
17: em\_hum:\*  
18: em\_in:\*  
19: em\_mu:\*  
20: em\_om:\*  
21: em\_or:\*  
22: em\_ov:\*  
23: em\_pat:\*  
24: em\_ph:\*  
25: em\_pl:\*  
26: em\_ro:\*  
27: em\_srs:\*  
28: em\_un:\*

29: em\_vl:\*  
30: em\_hcg\_hum:\*  
31: em\_hcg\_inv:\*  
32: em\_hcg\_other:\*  
33: em\_hcg\_mus:\*  
34: em\_hcg\_pln:\*  
35: em\_hcg\_rtd:\*  
36: em\_hcg\_mam:\*  
37: em\_hcg\_vrt:\*  
38: em\_sy:\*  
39: em\_hcgo\_hum:\*  
40: em\_hcgo\_mus:\*  
41: em\_hcgo\_other:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	552	100.0	1152	3 SCWCPROT	J03946 S.mansoni c
2	552	100.0	1477	3 SMHPIELAS	Z70296 S.mansoni e
3	519	94.0	3045	3 SMU31768	U31768 Schistosoma
4	503	91.1	943	3 AFS10341	AFS10341 Schistosoma
5	479	86.8	904	3 AFS10339	AFS10339 Schistosoma
6	458	83.0	419	3 AFS10342	AFS10342 Schistosoma
7	442	80.1	1163	3 SMU31769	U31769 Schistosoma
8	433	78.4	431	3 AFS10343	AFS10343 Schistosoma
9	267	48.4	886	3 AFS10340	AFS10340 Schistosoma
10	119	21.6	1479	3 AY061167	AY061167 Drosophila
11	119	21.6	4676	2 AC017985	AC017985 Drosophila
12	119	21.6	19947	3 AE002960	AE002960 Drosophila
13	112.5	20.4	198282	4 AC007415	AC007415 Drosophila
14	108.5	19.7	929	2 BTRYRPTMR	X94982 B.taurus mr
15	105	19.0	2197	4 M33494	M33494 Homo sapien
16	105	19.0	2280	9 AF099143	AF099143 Homo sapi
17	105	19.0	5456	9 AF099144	AF099144 Homo sapi
18	105	19.0	30228	9 HS030A1	AL031704 Human DNA
19	105	19.0	36006	9 HS030A1	AL031704 Human DNA
20	105	19.0	195680	9 AC120498	AC120498 Homo sapi
21	105	19.0	265786	9 AE006466	AE006466 Homo sapi
22	102.5	18.6	1002	4 OAR18223	Y18223 Ovis aries
23	102.5	18.6	3516	10 AF361940	AF361940 Mus muscu
24	102	18.5	944	10 MNRESPV	X78542 M.musculus
25	102	18.5	1090	10 MUSMACP	I21853 Mus muscu
26	102	18.5	1102	10 BC024374	BC024374 Mus muscu
27	102	18.5	1108	6 AR080464	AR080464 Sequence
28	102	18.5	1108	10 MUSPROTGB	MS7626 Mouse prote
29	102	18.5	2218	6 AR112352	AR112352 Sequence
30	102	18.5	2218	6 AR163667	AR163667 Sequence
31	102	18.5	2259	6 AR112351	AR112351 Sequence
32	102	18.5	2259	6 AR163666	AR163666 Sequence
33	102	18.5	3757	6 AR080463	AR080463 Sequence
34	102	18.5	3757	10 MUSPROTGA	MS7625 Mouse prote
35	101.5	18.4	1455	9 AF099147	AF099147 Homo sapi
36	101.5	18.4	2575	9 AF098327	AF098327 Homo sapi
37	101.5	18.4	2955	9 AF318074	AF318074 Homo sapi
38	101.5	18.4	30228	9 HS303A1	AL031704 Human DNA
39	101.5	18.4	265786	9 AE006466	AE006466 Homo sapi
40	101	18.3	1219	6 AR080457	AR080457 Sequence
41	101	18.3	1219	10 MUMICT	D31789 Mongolian g
42	100	18.1	768	9 AF206666	AF206666 Homo sapi
43	100	18.1	884	9 AF206665	AF206665 Homo sapi
44	100	18.1	1154	6 AR080459	AR080459 Sequence
45	100	18.1	1154	9 HUMTRY	M30038 Human trypt

RESULT 1

ALIGNMENTS

[illegible]

Db	789	GGTGAATATTGAGAA	806
<hr/>			
RESULT 2			
SMHPIELAS	SMHPIELAS	1477 bp	DNA linear INV 29-MAR-1996
LOCUS	S. mansoni elastase Hpl gene.		
DEFINITION	5. mansoni		
ACCESSION	Z70296		
VERSION	Z70296.1		
KEYWORDS	GI:1240057		
SOURCE	elastase.		
ORGANISM	Schistosoma mansoni.		
	Schistosoma mansoni		
	Eukaryota; Metazoa; Platyhelminthes; Trematoda; Digenea;		
	Strigoidae; Schistosomatidae; Schistosomidae; Schistosoma.		
REFERENCE	1 (bases 1 to 1477)		
AUTHORS	Sayers,J.R., Price,H.P. and Doenhoff,M.J.		
JOURNAL	Unpublished		
REFERENCE	2 (bases 1 to 1477)		
AUTHORS	Sayers,J.R.		
TITLE	Direct Submission		
JOURNAL	Submitted (21-MAR-1996) Jon R Sayers, Medicine and Pharmacology, University of Sheffield, Royal Hallamshire Hospital, Sheffield, South Yorkshire, S10 2JF, UK		
FEATURES	Location/Qualifiers		
source	1..1477		
	/organism="Schistosoma mansoni"		
	/db_xref="taxon:6183"		
exon	16..168		
	/number=1		
CDS	join(88..168,310..627,1235..1477)		
	/note="consensus sequence of 3 PCR clones from genomic DNA"		
	/codon_start=1		
	/product="cercarial elastase HPl"		
	/protein_id="CAA9312.1"		
	/db_xref="GI:1240058"		
	/translation="MCTGLSVSTRVAVLTAGHCVCPLPVTRVSFLTLRNGDOQIHQ PSGVKVAGYWPCMSARORBPJAOTLSEGDIALVMIAQWNIOSGRIVSLPSPDI PPPTGGFTIYGEGRDDNDPRDPSRKNGITLKKGRTIMECHARNGNPYCKAQNFEQ LPKAQSGGPLPBLPSQFPVLGVVSHVETLENLPDIIIVEVASVARMDPVRNSI"		
intron	169..309		
	/number=1		
exon	310..627		
	/number=2		
intron	628..1234		
	/number=2		
exon	1235..1477		
	/number=3		
BASE COUNT	347 a 261 c 404 g 465 t		
ORIGIN			
Alignment Scores:			
Pred. No.:	2.44e-46	Length:	1477
Percent:	552.00	Matches:	106
Score Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	100.00%	Indels:	0
DB:	3	Gaps:	0
<hr/>			
US-10-020-441-2 (1-106) x SMHPIELAS (1-1477)			
Oy	1 ValSerPheLeuThrLeuArgAsnGlyAspGlnGnglylleHisHicGlnProSerGly	20	
Db	310 GTTTCATTCTTCACACTGAGATGCCAACCAAGGCATCCATCACCACCGCTTGGA	369	
Oy	21 VALLVSVALALAPROGLTYTymeProserCySmetsEratlaaysgLnArAyProille	40	
Db	370 GTTAAGGGTGCACCGAGATCATGCCCCCTTGATGTGGCACGACGACGAGGACCAATC	429	
Oy	41 AlaGInThrLeuSerGlyPheAspIleAlaIleValMetLeuAlaGlnMetValasndeu	60	
Db	430 GCACGACACTCAGTGGCATTCGATTTCGATTGTGAATGCTGGCTCAATGTGCAACTTA	489	

QY 61 GlnSerGlyIleArgValIleSerLeuProGlnProSerAspIleProProGlyIthr 80  
DB 490 CAGAGTGGATCATCAGATGATCAGTCTGCCACAGCCATCGATATCCGCCACCTGGA 549  
QY 81 GYValPheIleValGlyIThrGlyArgAspAspAspAspAspProSerArgIleAsn 100  
DB 550 GGATTTTTCATTGTTGTTATGAGAGGATGATACGACCGTATCCGTCACTGAAGAT 609  
QY 101 GYGLYIleLeuLysLys 106  
DB 610 GGTGAATTTTGAAGAA 627  
RESULT 3  
SMU31768 3045 bp DNA linear INV 28-JUN-1996  
LOCUS Schistosoma mansoni elastase gene, 3045 bp clone, complete cds.  
DEFINITION U31768  
ACCESSION U31768  
VERSION U31768.1 GI:1103828  
KEYWORDS  
SOURCE blood fluke.  
ORGANISM Schistosoma mansoni  
Eukaryota; Metazoa; Platyhelminthes; Trematoda; Digenea;  
Strigeidida; Schistosomatidae; Schistosomatidae; Schistosoma.  
REFERENCE 1 (bases 1 to 3045)  
AUTHORS Pierrot,C., Capron,A. and Khalife,J.  
TITLE Cloning and characterization of two genes encoding Schistosoma mansoni elastase  
JOURNOL Mol. Biochem. Parasitol. 75 (1), 113-117 (1995)  
MEDLINE 96362066  
PUBMED 8720180  
REFERENCE 2 (bases 1 to 3045)  
AUTHORS Pierrot,C.  
TITLE Direct Submission  
JOURNOL Submitted (18-JUL-1995) Christine Pierrot, CIBP, INSERM U 167 - Institut Pasteur de Lille, 1, rue du Professeur Calmette, Lille 59019, France  
FEATURES  
source  
1..3045  
/organism="Schistosoma mansoni"  
/strain="Puerto rico"  
/db\_xref="taxon:6183"  
join(1..1465,1588..1935,2525..2910)  
/product="elastase"  
1..1231  
1153..1160  
1162..1167  
join(1232..1465,1588..1935,2525..2767)  
/codon\_start=1  
/product="elastase"  
/protein\_id="AAC46967.1"  
/db\_xref="GI:1103829"  
/translation="MSNRKRFVLVTLTFTYCTIFERVSTWLIRSGEPVQHRTEPPIAF  
LTERMTCTSLVSTRVLTAHGCVCSPLPIRVLCLEQVSPFLTRNGDQGIHQPS  
GKVAPEGMPSCMSARGRPIAQTLSGPDIAIVMLAQVNLQSGITVSLPOASDIPR  
PGTGVFIVIGYGRDNDNRDPSRNGGILKKKGILVGRATIMECRHATNGNPICVAGGN  
FGOLPAPDSDSGPPLPSQGPVLGVSHGVTLPNLPDIIIVEYASVARNLDFVRNIT"  
3'UTR  
polyA\_signal  
2886..2891  
polyA\_signal  
2906..2910  
BASE COUNT 851 a 579 c 691 g 924 t  
ORIGIN  
Alignment Scores:  
Pred. No.: 1.05e-42 Length: 3045  
Score: 519.00 Matches: 101  
Percent Similarity: 96.23% Conservative: 1  
Best Local Similarity: 95.28% Mismatches: 4  
Query Match: 94.02% Indels: 0  
Gaps: 0  
US-10-020-441-2 (1-106) x SMU31768 (1-3045)  
QY 1 ValSerPheLeuThrLeuArgAsnGlyAspGlnGlnGlyIleHisIleGlnProSerGly 20

DB 1603 GTTTCGTTTCTCACACTGAGAAACGGCGACCAAGGCATCATCACCAACATCTGGA 1662  
QY 21 ValLysValAlaProGlyIThrMetProSerCysMetSerAlaArgGlnArgProIle 40  
DB 1663 GTTAAAGTGGCACAGAGATATATGCCCTTTTATGATCGGACGAAGAGGAGACCAATC 1722  
QY 41 AlaGlnThrLeuSerGlyPheAspIleAlaIleValIleMetLeuAlaGlnMetValAsnLeu 60  
DB 1723 GCACAAACACTCAGTGGATTCACATTGGCAATTGTAACTCGCTCAATGGTCAACTTA 1782  
QY 61 GlnSerGlyIleArgValIleSerLeuProGlnProSerAspIleProProGlyIthr 80  
DB 1783 CAGAGTGGATCATCAGATGATCAGTCTGCCACAGCCATCGATATACCGACACCTGGA 1842  
QY 81 GYValPheIleValGlyIThrGlyArgAspAspAspAspAspProSerArgIleAsn 100  
DB 1843 GGATTTTTCATTGTTGTTATGAGAGGATGATACGACCGTATCCGTCTGTAAAT 1902  
QY 101 GYGLYIleLeuLysLys 106  
DB 1903 GGTGAATTTTGAAGAA 1920  
RESULT 4  
AF510341 943 bp mRNA linear INV 01-JUL-2002  
LOCUS Schistosoma haematobium elastase 1a (CE-1a) mRNA, complete cds.  
DEFINITION AF510341  
ACCESSION AF510341  
VERSION AF510341.1 GI:21217534  
KEYWORDS  
SOURCE Schistosoma haematobium.  
ORGANISM Schistosoma haematobium.  
Eukaryota; Metazoa; Platyhelminthes; Trematoda; Digenea;  
Strigeidida; Schistosomatidae; Schistosomatidae; Schistosoma.  
REFERENCE 1 (bases 1 to 943)  
AUTHORS Salter,J.P., Choe,Y., Albrecht,H., Franklin,C., Lim,K.C., Craik,C.S., and McKerrow,J.H.  
TITLE Cercarial elastase is encoded by a functionally conserved gene family across multiple species of schistosomes  
JOURNOL J. Biol. Chem. 277 (27), 24618-24624 (2002)  
MEDLINE 22086216  
PUBMED 11986325  
REFERENCE 2 (bases 1 to 943)  
AUTHORS Salter,J.P., Choe,Y., Albrecht,H., Franklin,C., Lim,K.C., Craik,C. and McKerrow,J.H.  
TITLE Direct Submission  
JOURNOL Submitted (08-MAY-2002) Pathology, University of California - San Francisco, 513 Parnassus, Box 0511, San Francisco, CA 94143, USA  
FEATURES  
source  
1..943  
/organism="Schistosoma haematobium"  
/db\_xref="taxon:6185"  
1..943  
/gene="CE-1a"  
/note="ShCE-1a"  
1..117  
/gene="CE-1a"  
18..809  
/gene="CE-1a"  
/note="Protease"  
/codon\_start=1  
/product="elastase 1a"  
/protein\_id="AA043943.1"  
/db\_xref="GI:21217535"  
/translation="MLNRKRFVLVTLTFTYCTIFERVSTWLIRSGEPVQHRTEPPIAF  
LTERMTCTSLVSTRVLTAHGCVCSPLPIRVLCLEQVSPFLTRNGDQGIHQPSGVA  
GYMPCMSARORRPIAQTLSGPDIAIVMLAQVNLQSGITVSLPOASDIPRPGTGV  
IVGGRDNDNRDPSRNGGILKKKGILVGRATIMECRHATNGNPICVSGQNFQGLPAPD  
SG GPPLPSQGPVLGVSHGVTLPNLPDIIIVEYASVARNLDFVRNIT"  
3'UTR  
polyA\_signal  
810..943  
/gene="CE-1a"  
919..924  
/gene="CE-1a"

BASE COUNT 263 a 220 c 210 g 250 t  
ORIGIN

Alignment Scores:  
Pred. No.: 1,31e-41 Length: 943  
Score: 503.00 Matches: 96  
Percent Similarity: 95.28% Conservative: 5  
Best Local Similarity: 90.57% Mismatches: 5  
Query Match: 91.12% Indels: 0  
DB: 3 Gaps: 0

US-10-020-441-2 (1-106) x AF510341 (1-943)

Qy 1 ValSerPheLeuThrLeuArgAsnGlyAspGlnGlnGlyIleHisGlnProSerGly 20  
Db 249 GTTTCATTTCTCAGACTGAGATGCGACCAACAGGATCATCATCAACCGTCTGA 308  
Qy 21 ValLysValAlaProGlyTyrMetProSerCysMetSerAlaArgGlnArgProIle 40  
Db 309 GTGGGTGTAGCACCAGATACATGCTCTTGTATGTCCGACGACAGAGAGACCAATC 368  
Qy 41 AlaGlnThrLeuSerGlyPheAspIleAlaIleValMetLeuAlaGlnMetValAsnLeu 60  
Db 369 CAACAGACACTCAGTGGATTGACATTCGCAATGTATCGCTCGCAATGTGCACTTA 428  
Qy 61 GlnSerGlyIleArgValIleSerLeuProGlnProSerAspIleProProGlyThr 80  
Db 429 CAGACTGGATACAGATGCTCAGTCTGCCACACGACGATTTACCGAGACTGAACT 488  
Qy 81 GlyValPheIleValGlyTyrGlyArgAspAspAspAspAspProSerArgLysAsn 100  
Db 489 CCGGTTTTCATTGTTGTTATGGAAGGATGATTAACGACCGTGCATCGTCACGCAAAAT 548  
Qy 101 GlyGlyIleLeuLysLys 106  
Db 549 GGTGGATTTTTAAGAA 566

RESULT 5  
AF510339 904 bp mRNA linear INV 01-JUL-2002  
LOCUS Schistosoma mansoni elastase 2a (CE-2a) mRNA, complete cds.  
DEFINITION AF510339  
ACCESSION AF510339.1 GI:21217530  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM Schistosoma mansoni.  
Eukaryota; Metazoa; Platyhelminthes; Trematoda; Digenea;  
Strigeidida; Schistosomatidae; Schistosomatidae; Schistosoma.  
REFERENCE 1 (bases 1 to 904)  
AUTHORS Salter,J.P., Choe,Y., Albrecht,H., Franklin,C., Lim,K.C.,  
Craik,C.S. and McKerrow,J.H.  
TITLE Cercarial elastase is encoded by a functionally conserved gene  
family across multiple species of schistosomes  
JOURNAL J. Biol. Chem. 277 (27), 24618-24624 (2002)  
MEDLINE 22086216  
PUBMED 11986325  
JOURNAL 2 (bases 1 to 904)  
PUBMED 11986325  
AUTHORS Salter,J.P., Choe,Y., Albrecht,H., Franklin,C., Lim,K.-C., Craik,C.  
and McKerrow,J.H.  
TITLE Direct Submission  
JOURNAL Submitted (08-MAY-2002) Pathology, University of California - San  
Francisco, 513 Parnassus, Box 0511, San Francisco, CA 94143, USA  
FEATURES  
source  
1..904  
Location/Qualifiers  
/organism="Schistosoma mansoni"  
/db\_xref="taxon:6183"  
1..904  
/gene="CE-2a"  
/note="SmCE-2a"  
1..792  
/gene="CE-2a"  
/note="protease"  
/codon\_start=1

/product="elastase 2a"  
/protein\_id="AA043941.1"  
/db\_xref="GI:21217531"  
/translation="MLNGRTFLMTLTFTYCLTFEVRVSTWLVRKGEVPQDRTEFPYIAF  
VRRPRMCTGSLVSTRAVLTAHGVCSPMVVOVSLTLTNGDQGIHDSGVKAP  
EYHPSCTASQRRIROTLGFDIATMLAQMNLOSGIIVISLRQASDIPTPTDVF  
IVGGRDNDPRSRAGGILKGRATVMECKSTSNPLCVQAIYFGQITAFGDSG  
GPLRSPQGPVLGVSHGVTLSNRLDVLVEVASVARMLGFSNNI"  
polyA\_signal  
872..877  
/gene="CE-2a"  
BASE COUNT 239 a 200 c 224 g 241 t  
ORIGIN

Alignment Scores:  
Pred. No.: 3,23e-39 Length: 904  
Score: 479.00 Matches: 94  
Percent Similarity: 91.51% Conservative: 3  
Best Local Similarity: 88.68% Mismatches: 9  
Query Match: 86.78% Indels: 0  
DB: 3 Gaps: 0

US-10-020-441-2 (1-106) x AF510339 (1-904)

Qy 1 ValSerPheLeuThrLeuArgAsnGlyAspGlnGlnGlyIleHisGlnProSerGly 20  
Db 232 GTTTCATTTCTCAGACTGAGATGCGACCAACAGGATCATCAACACCGTCTGA 291  
Qy 21 ValLysValAlaProGlyTyrMetProSerCysMetSerAlaArgGlnArgProIle 40  
Db 232 GTTAAGTACGCCAAGATACCTGCTTGTACGGGTCCGACAGAGAGAAATC 351  
Qy 41 AlaGlnThrLeuSerGlyPheAspIleAlaIleValMetLeuAlaGlnMetValAsnLeu 60  
Db 352 AGACAGACACTCAGTGGATTGACATTCGCAATGATGCTCGCAATGGTCAACTTA 411  
Qy 61 GlnSerGlyIleArgValIleSerLeuProGlnProSerAspIleProProGlyThr 80  
Db 412 CAGATGGATCAGATGATCAGTCTGCCACAGGATCGATATACGACACCTGAATC 471  
Qy 81 GlyValPheIleValGlyTyrGlyArgAspAspAspAspAspProSerArgLysAsn 100  
Db 472 GATGTTTTCATTGTTGTTATGGAAGGATGATTAACGACCGTGCATCGTGGTAGACT 531  
Qy 101 GlyGlyIleLeuLysLys 106  
Db 532 GGTGGATTTTGAAGAA 549

RESULT 6  
AF510342 419 bp mRNA linear INV 01-JUL-2002  
LOCUS Schistosoma haematobium elastase 1b (CE-1b) mRNA, partial cds.  
DEFINITION AF510342  
ACCESSION AF510342.1 GI:21217536  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM Schistosoma haematobium.  
Eukaryota; Metazoa; Platyhelminthes; Trematoda; Digenea;  
Strigeidida; Schistosomatidae; Schistosomatidae; Schistosoma.  
REFERENCE 1 (bases 1 to 419)  
AUTHORS Salter,J.P., Choe,Y., Albrecht,H., Franklin,C., Lim,K.C.,  
Craik,C.S. and McKerrow,J.H.  
TITLE Cercarial elastase is encoded by a functionally conserved gene  
family across multiple species of schistosomes  
JOURNAL J. Biol. Chem. 277 (27), 24618-24624 (2002)  
MEDLINE 22086216  
PUBMED 11986325  
JOURNAL 2 (bases 1 to 419)  
PUBMED 11986325  
AUTHORS Salter,J.P., Choe,Y., Albrecht,H., Franklin,C., Lim,K.-C., Craik,C.  
and McKerrow,J.H.  
TITLE Direct Submission  
JOURNAL Submitted (08-MAY-2002) Pathology, University of California - San  
Francisco, 513 Parnassus, Box 0511, San Francisco, CA 94143, USA  
FEATURES  
Location/Qualifiers

```
source
1. .419
/organism="Schistosoma haematobium"
/db_xref="taxon:6185"
<1. .>419
/gene="CE-1b"
/ncbi="ShCE-1b"
<1. .>419
/gene="CE-1b"
/ncbi="protease"
/codon_start=1
/product="elastase 1b"
/protein_id="AA043944.1"
/db_xref="GI:21217537"
/translation="DQCGIHQPSGVVAPGMPSCMSARORPIKOTLSGFDIATVS
LAQVNLQGIKRLSLPQPTDIPRGTFVIGYGRDNDNDPQKNGGILKKGRATL
MECHHTMGDILCYKSGNFGOLPAPRDSGGLPSP"

BASE COUNT      118 a      104 c      102 g      95 t
ORIGIN

Alignment Scores:
Pred. No.:      1.92e-37      Length:      419
Score:      458.00      Matches:      87
Percent Similarity:      93.81%      Conservative:      4
Best Local Similarity:      89.69%      Mismatches:      6
Query Match:      82.97%      Indels:      0
DB:      3      Gaps:      0

US-10-020-441-2 (1-106) x AF510342 (1-419)

QY 10 AapGInGInGlyIIEHIEHISGInProSeRGlyVAllyVAlaIaEProGlyTYrMePro 29
Db 1 GACCAACAAGGCATCATCACCAACGCTCTGGAGTGTGTGACACCAAGATATATGCCC 60
QY 30 SerCyMeSeSerAlaArgGlnArgArgProIleAlaGlnThrLeuSerGlyPheAspIle 49
Db 61 TCTGTATGTCTGGCAGACAGAGAGAACCAATCAACACACTCATGTGATTCGACATT 120
QY 50 AlaIleValMetLeuAlaGlnMetValaenLeuGlnSerGlyIleArgValIleSerIleu 69
Db 121 GCAATTGTAATCGCTCGCTCAATTGTCACACTTACAGACTCGAATCGAGTCTCACTTG 180
QY 70 ProGlnProSeSerAspIleProProGlyIleArgValaPheIleValGlyTYrGlyArg 89
Db 181 CCACAGCCCAACGATATACCGAGACTGTGAACTCCGGTTTCATTGTTGTTATGGAAG 240
QY 90 AspAraPheAspArgAspProSeSerArgLySaenGlyGlyIleLeuLeuLys 106
Db 241 GATGATTAACGACCGATCCACCAACGTAAATAATGTGGAAATTTTAAAGAA 291

RESULT 7
SMU31769      1163 bp      DNA      linear      INV 28-JAN-1996
LOCUS      Schistosoma mansoni elastase gene, 1163 bp clone, complete cds.
DEFINITION      U31769
ACCESSION      U31769
VERSION      U31769.1 GI:1103830
KEYWORDS
SOURCE
ORGANISM      blood fluke.
Schistosoma mansoni
Eukaryota; Metazoa; Platyhelminthes; Trematoda; Digenea;
Strigoida; Schistosomatidae; Schistosomatidae; Schistosoma.
1 (bases 1 to 1163)
Pierrot,C., Capron,A. and Khalife,J.
Cloning and characterization of two genes encoding Schistosoma
mansoni elastase
Mol. Biochem. Parasitol. 75 (1), 113-117 (1995)
JOURNAL
MEDLINE      8720180
PUBMED      8720180
REFERENCE
AUTHORS      Pierrot,C.
TITLE      Direct Submision
JOURNAL      Submitted (18-JUL-1995) Christine Pierrot, CIBP, INSERM U167 -
INstitut Pasteur de Lille, 1, rue du Professeur Calmette, Lille,
59019 France
```

```
FEATURES
source
Location/Qualifiers
1. .1163
/organism="Schistosoma mansoni"
/strain="Puerto rico"
/db_xref="taxon:6183"
join(<1. .234,360. .668,933. .>1163)
join(1. .234,360. .668,933. .1163)
/codon_start=1
/product="elastase"
/protein_id="AAC46968.1"
/db_xref="GI:1103831"
/translation="MSNRMRPLVLTFLTCLEFRVSTWLTIRSDQPVQKTEPPFIAY
IASKSMCTGSLVSTRVLTAGHCVCPEMPYIKQVYFLTRNDQGIHQPSGVKY
AFEXMPSCTASRORRRIQRTLSGPAITAMLAQVNLQGIKRLSLPQPTDIPRGTD
VFIYGRDNDNDPDRRAGGRATVTECHRETHVNPICVAGPNSGQLGPDSGGP
LHPDQGVLCVSHGVTLSLPEYVAVRLNVRNLI"

BASE COUNT      311 a      226 c      275 g      351 t
ORIGIN

Alignment Scores:
Pred. No.:      2.18e-35      Length:      1163
Score:      442.00      Matches:      86
Percent Similarity:      88.24%      Conservative:      4
Best Local Similarity:      84.31%      Mismatches:      12
Query Match:      80.07%      Indels:      0
DB:      3      Gaps:      0

US-10-020-441-2 (1-106) x SMU31769 (1-1163)

QY 1 ValSerPheLeuThrLeuArgAsnGlyAapGInGInGlyIIEHIEHISGInProSeRGly 20
Db 363 GTTACATTTCTCATCTTGGAGATGGAGCAACCAAGGCATCATCACACCGTCTGGA 422
QY 21 ValIyVAlaIaEProGlyTYrMeProSeRGlyMeSeSerAlaArgGlnArgArgProIle 40
Db 423 GTTAAAGTAAACCAAGATTAATGCTCTGTAGCGGTCAACAGAGAGGAAGAAATC 482
QY 41 AlaGlnThrLeuSerGlyPheAspIleAlaIleValMetLeuAlaGlnMetValaenLeu 60
Db 483 AGACAGACACTCAGTGAATTCGCCATTGCACACTGCATGCTCCAAATGCTCAACTTA 542
QY 61 GlnSerGlyIleArgValIleSerLeuProGlnProSeSerAspIleProProGlyIleThr 80
Db 543 CAGATGTGAATCAGATGATGTTGCTGCCACAGGATTCGATATACGACACTGGAATC 602
QY 81 GlyValPheIleValGlyTYrGlyArgAspAraPheAspProSeSerArgLySaen 100
Db 603 GATGTTTCATTTGTTGTTATGGAAGGATATACGACCGGTGATCCGCTGAGAGCT 662
QY 101 GIGlyI 102
Db 663 GGTGGA 668

RESULT 8
AF510343      431 bp      mRNA      linear      INV 01-JUN-2002
LOCUS      Schistosomacium douthitti elastase 1a (CE-1a) mRNA, partial cds.
DEFINITION      AF510343
ACCESSION      AF510343
VERSION      AF510343.1 GI:21217538
KEYWORDS
SOURCE
ORGANISM      Schistosomacium douthitti.
Schistosomacium douthitti.
Eukaryota; Metazoa; Platyhelminthes; Trematoda; Digenea;
Strigoida; Schistosomatidae; Schistosomatidae; Schistosomacium.
1 (bases 1 to 431)
Salter,J.P., Choe,Y., Albrecht,H., Franklin,C., Lim,K.C.,
Craik,C.S. and McKerrow,J.H.
Cercarial elastase is encoded by a functionally conserved gene
family across multiple species of schistosomes
J. Biol. Chem. 277 (27), 24618-24624 (2002)
JOURNAL
MEDLINE      22086216
PUBMED      11986325
REFERENCE
AUTHORS      Salter,J.P., Choe,Y., Albrecht,H., Franklin,C., Lim,K.C.,
Craik,C.S. and McKerrow,J.H.
TITLE      family across multiple species of schistosomes
JOURNAL      Submitted (18-JUL-1995) Christine Pierrot, CIBP, INSERM U167 -
INstitut Pasteur de Lille, 1, rue du Professeur Calmette, Lille,
59019 France
```

AUTHORS Salter, J. P., Choe, Y., Albrecht, H., Franklin, C., Lim, K.-C., Craik, C. and McKerrow, J. H.  
TITLE Direct Submission  
JOURNAL Submitted (08-MAY-2002) Pathology, University of California - San Francisco, 513 Parnassus, Box 0511, San Francisco, CA 94143, USA  
FEATURES Location/Qualifiers  
source 1..431  
/organism="Schistosoma matsumotii"  
/db\_xref="taxon:38741"  
gene <1..>431  
/gene="CE-1a"  
/note="ScdC-1a"  
CDS <1..>431  
/gene="CE-1a"  
/note="protease"  
/codon\_start=1  
/product="elastase 1a"  
/protein\_id="AA043945.1"  
/db\_xref="GI:21217533"  
/translation="DQGIHQPSGVVAPEYMSQSSRRQRRVQTLTSGFDIAIVL  
LAEMVNLQGIKVLSPQTDIPAGTGVFVIGRDNDPSRBNGLIKGRATV  
MESKSTGNRI CVQGRVFPQIPAPDGSGLPLSPQNSIS"  
BASE COUNT 117 a 108 c 109 g 97 t  
ORIGIN  
Alignment Scores:  
Pred. No.: 6.43e-35 Length: 431  
Score: 433.00 Matches: 81  
Percent Similarity: 92.78% Conservative: 9  
Best Local Similarity: 83.51% Mismatches: 7  
Query Match: 78.44% Indels: 0  
DB: 3 Gaps: 0  
US-10-020-441-2 (1-106) x AF510343 (1-431)  
QY 10 AAGGNGNGLYIIEHISGINSERGLYVALYVAIALAProGLYTYrMetPro 29  
Db 1 GACCAACAGGATCATCACCAACGCTCGAGTGTGTAGACACAGAAATACATGCCA 60  
QY 30 SArCyMeSerAlaArgInArgProIleAlaGlnThleuSerGlyPheAspIle 49  
Db 61 TCTTGCGCTCGCGACGACAGAGAGACGATCAACACACTCGATGATTCGATATT 120  
QY 50 AAlaIleValMetLeuAlaGlnMetValAsnLeuGlnSerGlyIleArgValIleSerLeu 69  
Db 121 GCATTTGATTTGCTGCTGAATGTCATTCACACCGGATCAAGGTGCTCACTCG 180  
QY 70 ProGlnProSerAspIleProProPogIlyThrGlyValPheIleValGlyTYrGlyArg 89  
Db 181 CCACAGCCACGATATACCGGACCTGGAACTCCCGTTTTCATTTGTTGTTATGGAAG 240  
QY 90 AAPAAsnAspArgAspProSerArgLyAsnGlyGlyIleLeuLysLys 106  
Db 241 GATGATACGACCGTATCCGTACCGACGAGAAATGGTGGATTTTGAAGAA 291  
RESULT 9  
AF510340 886 bp mRNA linear INV 01-JUL-2002  
LOCUS Schistosoma mansoni elastase 2b (CE-2b) mRNA, complete cds.  
ACCESSION AF510340  
VERSION AF510340.1 GI:21217532  
KEYWORDS Schistosoma mansoni.  
ORGANISM Schistosoma  
Eukaryota; Metazoa; Platyhelminthes; Trematoda; Digenea;  
Strigeidida; Schistosomatidae; Schistosomatidae; Schistosoma.  
REFERENCE 1 (bases 1 to 886)  
AUTHORS Salter, J. P., Choe, Y., Albrecht, H., Franklin, C., Lim, K.-C.,  
Craik, C. S. and McKerrow, J. H.  
TITLE Cercarial elastase is encoded by a functionally conserved gene  
family across multiple species of schistosomes  
JOURNAL J. Biol. Chem. 277 (27), 24618-24624 (2002)  
MEDLINE 22086216

PUBMED 11986325  
REFERENCE 2 (bases 1 to 886)  
AUTHORS Salter, J. P., Choe, Y., Albrecht, H., Franklin, C., Lim, K.-C., Craik, C. and McKerrow, J. H.  
TITLE Direct Submission  
JOURNAL Submitted (08-MAY-2002) Pathology, University of California - San Francisco, 513 Parnassus, Box 0511, San Francisco, CA 94143, USA  
FEATURES Location/Qualifiers  
source 1..886  
/organism="Schistosoma mansoni"  
/db\_xref="taxon:6183"  
gene <1..>886  
/gene="CE-2b"  
/note="SmCE-2b"  
CDS <1..>792  
/gene="CE-2b"  
/note="protease"  
/codon\_start=1  
/product="elastase 2b"  
/protein\_id="AA043942.1"  
/db\_xref="GI:21217533"  
/translation="MLNGMTFLIVTLFTYVLTCCVSTWILRSGEVQQRTEPFIAL  
LMDASMCTGSIVSRAVLATAGHCQTPVRSFSLSEFDORTINHRLEIKVAP  
EYNVCCQLKRNKRITKSLGSDMAITITLVNVLKRGVYISLAELDIPESIAV  
WVGYGQDIRDPDPSGRYGLIKKSGATITMCRHRTDPDPCVRFQPNKQIAGPDSG  
GPLLTLPQGPVIVASNGVFLPALADLCEVRSVPRMLKFLIPNI"  
BASE COUNT 257 a 165 c 191 g 273 t  
ORIGIN  
polyA\_signal  
Alignment Scores:  
Pred. No.: 6.42e-18 Length: 886  
Score: 267.00 Matches: 52  
Percent Similarity: 69.81% Conservative: 22  
Best Local Similarity: 49.06% Mismatches: 32  
Query Match: 48.37% Indels: 0  
DB: 3 Gaps: 0  
US-10-020-441-2 (1-106) x AF510340 (1-886)  
QY 1 ValSerPheLeuThrLeuArgAsnGlyAspGlnGlnGlyIIEHISGINSERGLY 20  
Db 232 GTTTCGTTTCATCTCAATGCAATTTGACGAGAACTATTATCATCGACCTTTGGA 291  
QY 21 ValLysValAlaProGlyTYrMetProSerCyMetSerAlaArgGlnArgProIle 40  
Db 292 ATTAAAGTTGCACGAGATATATCTGTTTTCAGCTGAAACGAGAAATTAAGCATT 351  
QY 41 AlaGlnThrLeuSerGlyPheAspIleAlaIleValMetLeuAlaGlnMetValAsnLeu 60  
Db 352 ACGAGTCACTTGGTGATATGATATGCAATTATTAACCTCAATTAATTGTCACCTG 411  
QY 61 GlnSerGlyIleArgValIleSerLeuProGlnProSerAspIleProProGlyThr 80  
Db 412 GAGACTGAGTTAAAGTATTAGTCTGGCAGCTGAATTGATATACCTATACCAAAAGT 471  
QY 81 GlyValPheIleValGlyTYrGlyArgAspAsnAspArgAspProSerArgLyAsn 100  
Db 472 ATTGCTTATATGATAGTATGATGACAGGATTTAGGATCTCGATTCATCCGCTGATAT 531  
QY 101 GLYGLYIleLeuLysLys 106  
Db 532 GGTGGAATTTGAAAAA 549  
RESULT 10  
AY061167  
LOCUS AY061167 1479 bp mRNA linear INV 08-NOV-2001  
DEFINITION Drosophila melanogaster LDI3269 full length cDNA.  
ACCESSION AY061167  
VERSION AY061167.1 GI:16768991  
KEYWORDS FLI cDNA.  
SOURCE Drosophila melanogaster.

FEATURES	location/Qualifiers
source	1..1479
	/organism="Drosophila melanogaster"
	/strain="Y; cn bw sp"
	/db_xref="taxon:7227"
CDS	93..1358

misc_feature	469 a	273 c	312 g	425 t
BASE COUNT				
ORIGIN				

Alignment Scores:	
Pred. No.:	0.00807
Score:	119.00
Percent Similarity:	48.42%
Best Local Similarity:	34.74%
Query Match:	21.56%
DB:	3
US-10-020-441-2 (1-106) × AY061167 (1-1479)	
	Length: 1479
	Matches: 33
	Conservative: 13
	Mismatches: 27
	Indels: 22
	Gaps: 4

```
OY      5  Thrlendrgnslvne-----Gingngllyllehsisglnpro 18
        |||::|||::|||::|||::|||::|||::|||::|||::|||::
Db      744  Acgtgcggacttgaaagcggaatcacaaaacatgaamgacctcttcaaggaa 800
        |||::|||::|||::|||::|||::|||::|||::|||::|||::
OY      19  Serclvalyvalia-----Proglytyrmetproserybmetserrlaarg 35
        |||::|||::|||::|||::|||::|||::|||::|||::
```

```

Db      804 AGATGGTTCAACGCTGTAATATGGACCCAGATTAT----- 839
QY      36 GlnAatGatgProIleIaIeGlnThreusercIyPheapIleIaIeValMetLeuIa 55
        |||||
Db      840 AACGAGMAAGCATGTGCT-----TATGATTTTCGCACCTGTGTAATCTAAGT 884
QY      56 GlnMetValaMetLeuGlnsercIyIleIaArgValIleIeserLeuProGlnProserAspIle 75
        |||||
Db      885 CAGCGCGTTACTTGGATGATGCATTAATGATGTAAATGTTAACTCAACCAAGATGATATT 944
QY      76 ProProGcgIyThrgIyValPheIleValGlyTyrgIyArgsp 90
        |||||
Db      945 CCGCAGCCCGGATACACATGTTTTTCCACCGCGCTGGGAAAAGAT 969

```

RESULT	11
LOCUS	AC017985
DEFINITION	AC017985 Drosophila melanogaster, 4676 bp DNA linear HTG 09-DEC-1999
ACCESSION	AC017985
VERSION	AC017985.1 GI:553205
KEYWORDS	HTG; HTGS PHASE2.
SOURCE	Drosophila melanogaster.
ORGANISM	Drosophila melanogaster.
REFERENCE	Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
AUTHORS	Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
TITLE	Ephydroidae; Drosophilidae; Drosophila.
JOURNAL	1 (bases 1 to 4676)
COMMENT	Adams,M. and Venter,J.C. Direct Submission Submitted (09-DEC-1999) Celera Genomics, 45 West Gude Drive, Rockville, MD, USA This sequence was identified as CDM:10212908 by the submitter.

```

FEATURES
  SOURCE
    1. .4676

```

```

/organism="Drosophila melanogaster"
/db_xref="taxon:7227"
BASE COUNT      1482 a      839 c      950 g      1405 t
ORIGIN
Alignment Scores:
Pred. No.:      0.0258      Length:      4676
Score:          119.00      Matches:     33
Percent Similarity: 48.42%      Conservative: 13
Best Local Similarity: 34.74%      Mismatches: 27
Query Match:    21.56%      Indels:     22
DB:             2          Gaps:         4

```

US-10-020-441-2 (1-106) X ACO17985 (1-4676)

Oy	5	ThruuLgaenGjYAsp-----	-GingGjYlIeIhEhI	GjIpro	18
		::: ::: ::: ::: :::	::: ::: ::: ::: :::		
Db	486	ACTGTCGAGCTGGAGAGTGGATACAAACAACAAATGAAGAAGCGCTTACCTATCAAGAA			545
Oy	19	SerGjYValIySvalAla-----	ProGjYIyrhMetProSerGyImetSerIlaIarg		35
		::: ::: ::: ::: :::	::: ::: ::: ::: :::		
Db	546	AGATCGGTTCAACTGTTATATTGGACCCCAACTTAT-----			581
Oy	36	GlnArgArgProIleAlaGlnThrLeuSerGjYpheaAspIleAlaIleValMetLeuAla			55
		::: ::: ::: ::: :::	::: ::: ::: ::: :::		
Db	582	AACGAGAGAACCAATAGCT-----	-TATGATTTTCGGACACTGTGTAATCTTAACT		626
Oy	56	GlnMetValaenLeuGlnSerGjYIleArgValIleSerLeuProGlnInProSerAspIle			75
		::: ::: ::: ::: :::	::: ::: ::: ::: :::		
Db	627	CAGCCGCTTACTTTGGATGATCACTTATATGTTTATATGTTTACTCTCAACAAGATGATATT			686
Oy	76	ProProProGjYIthrGjYValpheaIleValGjYIyrGjYIargAsp			90
		::: ::: ::: ::: :::	::: ::: ::: ::: :::		

Db 687 CCGCAGCCCGTACACATGTTTTCACCGCGTGGGAAAAGAT 731

RESULT 12  
AE002960/c 19947 bp DNA linear INV 06-OCT-2000  
LOCUS Drosophila melanogaster genomic scaffold 14200001385601, complete  
DEFINITION  
sequence.  
ACCESSION  
AE002960  
VERSION  
AE002960  
KEYWORDS  
GI:10729553  
SOURCE  
HTG.  
ORGANISM  
Drosophila melanogaster.  
Drosophila melanogaster.  
Eukaryota, Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;  
Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
Ephydroidea; Drosophilidae; Drosophila.  
REFERENCE  
1 (bases 1 to 19947)  
Adams,M.D., Celniker,S.E., Holt,R.A., Evans,C.A., Gocayne,J.D.,  
Amanatides,P.G., Scherer,S.E., Li,P.W., Hoskins,R.A., Galle,R.F.,  
George,R.A., Lewis,S.E., Richards,S., Ashburner,M., Henderson,S.N.,  
Sutton,G.G., Wortman,J.R., Yandell,M.D., Zhang,Q., Chen,L.X.,  
Brandon,R.C., Rogers,Y.H., Blazey,R.G., Champe,M., Pfeiffer,B.D.,  
Wan,K.H., Doyle,C., Baxter,E.G., Helt,G., Nelson,C.R., Gabor  
Milos,G.L., Abell,J.F., Agbayani,A., An,H.J., Basu,A.,  
Bardales,J., Bayraktaroglu,L., Beasley,E.M., Beeson,K.Y.,  
Bencos,P.V., Berman,B.P., Bhandari,D., Bolashakov,S., Borkova,D.,  
Borochan,M.R., Bouck,J., Brockstein,P., Brotlier,P., Burtis,K.C.,  
Busam,D.A., Butler,H., Cadieu,E., Center,A., Chandra,I.,  
Cheriy,J.M., Cawley,S., Dahlke,C., Davenport,L.B., Davies,P., de  
Pallos,B., Delcher,A., Deng,Z., Mays,A.D., Dew,I., Dietz,S.M.,  
Dodson,K., Doup,L.E., Downes,M., Dugan-Rocha,S., Dunkov,B.C.,  
Dunn,P., Durbin,K.J., Evangelista,C.C., Ferraz,C., Ferriera,S.,  
Flischiemann,W., Foster,C., Gabrielian,A.E., Garg,N.S.,  
Gelbart,W.M., Glasser,K., Glodek,A., Gong,F., Gorell,J.H., Gu,Z.,  
Guan,P., Harris,M., Harris,N.L., Harvey,D., Helman,T.J.,  
Hernandez,J.R., Houck,J., Hostin,D., Houston,K.A., Howland,T.J.,  
Wei,M.H., Ibegwan,C., Jalali,M., Kalush,B.E., Kodira,C.D., Kraft,C.,  
Kraut,S., Kulp,D., Lai,Z., Lasko,P., Lei,Y., Levitsky,A.A.,  
Li,J., Li,Z., Liang,Y., Lin,X., Liu,X., Matrei,B., McIntosh,T.C.,  
McLeod,M.P., McPherson,D., Meikunov,G., Milshina,N.V., Mobarry,C.,  
Mortis,J., Moshrefi,A., Mount,S.M., Moy,M., Murphy,B., Murphy,L.,  
Murphy,D.M., Nelson,D.L., Nelson,D.R., Nelson,K.A., Nixon,K.,  
Nusken,D.R., Pacle,J.M., Palazolo,M., Pittman,G.S., Pan,S.,  
Pollard,J., Puri,V., Reese,M.G., Reinert,K., Ramington,K.,  
Saunders,R.D., Scheeler,F., Shen,H., Shie,B.C., Siden-Kiamos,I.,  
Simpson,M., Skupski,M.P., Smith,T., Spier,E., Spradling,A.C.,  
Stapleton,M., Strong,R., Sun,E., Svitskas,R., Tector,C., Turner,R.,  
Venter,E., Wang,A.H., Wang,X., Wang,Z.Y., Wassarman,D.A.,  
Weinstock,G.M., Weisenbach,J., Williams,S.M., Woodage,T.,  
Worley,K.C., Wu,D., Yang,S., Yao,Q.A., Ye,J., Yeh,R.F.,  
Zaveri,J.S., Zhan,M., Zhang,G., Zhao,Q., Zheng,L., Zheng,X.H.,  
Zhong,F.N., Zhong,W., Zhou,X., Zhu,S., Zhu,X., Smith,H.O.,  
Gibbs,R.A., Myers,E.W., Rubin,G.M. and Venter,J.C.  
The genome sequence of *Drosophila melanogaster*  
Science 287 (5461), 2185-2195 (2000)  
20196006  
10731132  
TITLE  
JOURNAL  
MEDLINE  
PUBMED  
REFERENCE  
AUTHORS  
JOURNAL  
TITLE  
2 (bases 1 to 19947)  
Adams,M.D., Celniker,S.E., Gibbs,R.A., Rubin,G.M. and Venter,J.C.  
Direct Submission  
Submitted (21-MAR-2000) Celera Genomics, 45 West Gude Drive,  
Rockville, MD, USA  
On Oct 9, 2000 this sequence version replaced gi:7289359.  
COMMENT  
FEATURES  
source  
1. 19947  
/organism="Drosophila melanogaster"  
/db\_xref="taxon:7227"  
/chromosome="2R"  
/map="41C-41D"  
BASE COUNT 2019 a 1310 c 1239 g 2079 t 13300 others  
ORIGIN

Alignment Scores:  
Pred. No.: 0.111  
Score: 119.00  
Percent Similarity: 48.42%  
Best Local Similarity: 34.74%  
Query Match: 21.56%  
DB: 3  
Gaps: 4  
US-10-020-441-2 (1-106) x AE002960 (1-19947)  
Oy 5 ThrleuAagaNGlyAap-----GInGInGlyleHleHleGInPro 18  
Db 12017 ACTGTCAGCTGAGAGATGGAGATACACAAACATGAAGAGCGCTTACCTATATAGAA 11958  
Oy 19 SerGlyVallyValala-----ProGlyTyrMetProSerCyMetSerAlaary 35  
Db 11957 AGATGGTTCAACGCTTATATGACCCACATATAT----- 11922  
Oy 36 GlnAgaTgProIleAlaGlnThrleuSerGlyPheapIleAlaIleValMetleuAla 55  
Db 11921 AACAGAGACAGATGCT-----TATGATTCGCACTGTGTAATCTTAAGT 11877  
Oy 56 GlnMetValaenleuGlnSerGlyIlearyVallyIleSerleuProGlnProSerAspIle 75  
Db 11876 CAGCGGTTACTTGGATGATCATCATTAATGTTATGTTATCTTACCTCAACAAAGATGATAT 11817  
Oy 76 PropProGlyThrGlyValPheIleValGlyTyrGlyArgAsp 90  
Db 11816 CCGCAGCCCGTACACATGTTTTCACCGCGTGGGAAAAGAT 11772  
RESULT 13  
AC007415  
LOCUS  
DEFINITION  
Drosophila melanogaster chromosome 2 clone BACR05424 (D603) RPCT-98  
05.A.34 map 41C-41D strain Y; cn bw sp. \*\*\* SEQUENCING IN PROGRESS  
\*\*\*, 103 unordered pieces.  
ACCESSION  
AC007415  
VERSION  
AC007415.11 GI:7259656  
KEYWORDS  
HTG; HTGS PHASE1.  
SOURCE  
ORGANISM  
Drosophila melanogaster.  
Drosophila melanogaster.  
Eukaryota, Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;  
Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
Ephydroidea; Drosophilidae; Drosophila.  
REFERENCE  
1 (bases 1 to 198282)  
Celniker,S.E., Agbayani,A., Arcaina,T.T., Baxter,E., Blazey,R.G.,  
Burenhoff,C., Champe,M., Chavez,C., Chew,M., Ciesiolka,L.,  
Doyle,C.M., Farfan,D.E., Galle,R., George,R.A., Harris,N.L.,  
Hinkle,A., Hoskins,R.A., Houston,K.A., Hummasti,S.R., Karra,K.,  
Keatney,L., Lee,B., Lewis,S., Li,P., Ling,H., Moshrefi,A.R.,  
Moshrefi,M., Nixon,K., Pacle,J.M., Park,S., Pfeiffer,B.,  
Richards,S., Sethi,H., Svitskas,R.R., Wan,K.H., Webster,D.,  
Woolley,P., Yang,S., Yee,M., Yu,C. and Rubin,G.M.  
Sequencing of *Drosophila melanogaster*  
Unpublished  
2 (bases 1 to 198282)  
Celniker,S.E., Agbayani,A., Arcaina,T.T., Baxter,E., Blazey,R.G.,  
Burenhoff,C., Champe,M., Chavez,C., Chew,M., Ciesiolka,L.,  
Doyle,C.M., Farfan,D.E., Galle,R., George,R.A., Harris,N.L.,  
Hoskins,R.A., Houston,K.A., Hummasti,S.R., Karra,K., Keatney,L.,  
Kim,E., Lee,B., Lewis,S., Li,P., Lomontan,M.A., Maeda,P.,  
Moshrefi,A.R., Moshrefi,M., Nixon,K., Pacle,J.M., Park,S.,  
Pfeiffer,B., Poon,L., Seguiria,A., Sethi,H., Shit,E.,  
Svitskas,R.R., Wan,K.H., Weinburg,T., Zhang,R., Zieran,L.L. and  
Rubin,G.M.  
Direct Submission  
Submitted (28-APR-1999) Drosophila Genome Center, Lawrence Berkeley  
Laboratory, MS 64-121, Berkeley, CA 94720, USA  
On Mar 17, 2000 this sequence version replaced gi:7159347.  
For further information about this sequence, including its location  
and relationship to other sequences, please visit our sequence  
archive Web site (<http://www.fruitfly.org/sequence/>) or send email  
to [bdgs@fruitfly.berkeley.edu](mailto:bdgs@fruitfly.berkeley.edu). All contigs in this submission meet



the following cutoffs: length >= 200 bases.

\* NOTE: This is a 'working draft' sequence. It currently consists of 103 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown. \* This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved.

```
1 926: contig of 926 bp in length
* 927 1006: gap of unknown length
* 1007 1301: contig of 295 bp in length
* 1302 1381: gap of unknown length
* 1382 1860: contig of 479 bp in length
* 1861 1940: gap of unknown length
* 1941 2403: contig of 463 bp in length
* 2404 2483: gap of unknown length
* 2484 2895: contig of 412 bp in length
* 2896 2975: gap of unknown length
* 2976 3478: contig of 503 bp in length
* 3479 3558: gap of unknown length
* 3559 4034: contig of 476 bp in length
* 4035 4114: gap of unknown length
* 4115 4788: contig of 674 bp in length
* 4789 4868: gap of unknown length
* 4869 5549: contig of 681 bp in length
* 5550 5630: gap of unknown length
* 5630 6444: contig of 814 bp in length
* 6444 6523: gap of unknown length
* 6524 7697: contig of 1174 bp in length
* 7698 7777: gap of unknown length
* 7778 9243: contig of 1466 bp in length
* 9244 9323: gap of unknown length
* 9324 12162: contig of 2839 bp in length
* 12163 12242: gap of unknown length
* 12243 14263: contig of 2021 bp in length
* 14264 14343: gap of unknown length
* 14344 16104: contig of 1761 bp in length
* 16105 16184: gap of unknown length
* 16185 18228: contig of 2044 bp in length
* 18229 18308: gap of unknown length
* 18309 20874: contig of 2566 bp in length
* 20875 20954: gap of unknown length
* 20955 24951: contig of 3997 bp in length
* 24952 25031: gap of unknown length
* 25032 28493: contig of 3462 bp in length
* 28494 28573: gap of unknown length
* 28574 31026: contig of 2453 bp in length
* 31027 31106: gap of unknown length
* 31107 35546: contig of 4440 bp in length
* 35547 35626: gap of unknown length
* 35627 40849: contig of 5223 bp in length
* 40850 40929: gap of unknown length
* 40930 48761: contig of 7832 bp in length
* 48762 48841: gap of unknown length
* 48842 52656: contig of 3815 bp in length
* 52657 52736: gap of unknown length
* 52737 58188: contig of 5452 bp in length
* 58189 58268: gap of unknown length
* 58269 62896: contig of 4628 bp in length
* 62897 62976: gap of unknown length
* 62977 72175: contig of 9199 bp in length
* 72176 72255: gap of unknown length
* 72256 86109: contig of 13854 bp in length
* 86110 86189: gap of unknown length
* 86190 94772: contig of 8583 bp in length
* 94773 94852: gap of unknown length
* 94853 112088: contig of 17236 bp in length
* 112089 112168: gap of unknown length
* 112169 128075: contig of 15807 bp in length
* 128076 128155: gap of unknown length
* 128156 146641: contig of 18486 bp in length
* 146642 146721: gap of unknown length
```

```
* 146722 147805: contig of 1084 bp in length
* 147806 147885: gap of unknown length
* 147886 148190: contig of 305 bp in length
* 148191 148270: gap of unknown length
* 148271 149254: contig of 984 bp in length
* 149255 149334: gap of unknown length
* 149335 150046: contig of 712 bp in length
* 150047 150126: gap of unknown length
* 150127 151028: contig of 902 bp in length
* 151029 151108: gap of unknown length
* 151109 151716: contig of 608 bp in length
* 151717 151796: gap of unknown length
* 151797 152455: contig of 659 bp in length
* 152456 152535: gap of unknown length
* 152536 153204: contig of 669 bp in length
* 153205 153284: gap of unknown length
* 153285 153938: contig of 654 bp in length
* 153939 154018: gap of unknown length
* 154019 154280: contig of 262 bp in length
* 154281 154360: gap of unknown length
* 154361 155656: contig of 1296 bp in length
* 155657 155736: gap of unknown length
* 155737 156328: contig of 552 bp in length
* 156329 156408: gap of unknown length
* 156409 157266: contig of 858 bp in length
* 157267 157346: gap of unknown length
* 157347 157827: contig of 481 bp in length
* 157828 157907: gap of unknown length
* 157908 158993: contig of 1086 bp in length
* 158994 159073: gap of unknown length
* 159074 159316: contig of 243 bp in length
* 159317 159396: gap of unknown length
* 159397 159756: contig of 360 bp in length
* 159757 159836: gap of unknown length
* 159837 160697: contig of 861 bp in length
* 160698 160776: gap of unknown length
* 160778 161496: contig of 719 bp in length
* 161497 161576: gap of unknown length
* 161577 162155: contig of 579 bp in length
* 162156 162235: gap of unknown length
* 162236 162542: contig of 307 bp in length
* 162543 162622: gap of unknown length
* 162623 162964: contig of 342 bp in length
* 162965 163044: gap of unknown length
* 163045 163544: contig of 500 bp in length
* 163545 163624: gap of unknown length
* 163625 165092: contig of 1468 bp in length
* 165093 165172: gap of unknown length
* 165173 165866: contig of 694 bp in length
* 165867 165946: gap of unknown length
* 165947 166183: contig of 237 bp in length
* 166184 166263: gap of unknown length
* 166264 167263: contig of 1000 bp in length
* 167264 167343: gap of unknown length
* 167344 167883: contig of 540 bp in length
* 167884 167963: gap of unknown length
* 167964 168338: contig of 375 bp in length
* 168339 168418: gap of unknown length
* 168419 168719: contig of 301 bp in length
* 168720 168799: gap of unknown length
* 168800 169480: contig of 681 bp in length
* 169481 169560: gap of unknown length
* 169561 170067: contig of 507 bp in length
* 170068 170147: gap of unknown length
* 170148 170904: contig of 757 bp in length
* 170905 170984: gap of unknown length
* 170985 171769: contig of 785 bp in length
* 171770 171849: gap of unknown length
* 171850 172398: contig of 549 bp in length
* 172399 172478: gap of unknown length
* 172479 173200: contig of 722 bp in length
* 173201 173280: gap of unknown length
* 173281 173833: contig of 553 bp in length
```

```
* 173834 173913: gap of unknown length
* 173914 174940: contig of 1027 bp in length
* 174941 175020: gap of unknown length
* 175021 175732: contig of 712 bp in length
* 175733 175812: gap of unknown length
* 175813 176895: contig of 1083 bp in length
* 176896 176975: gap of unknown length
* 176976 177582: contig of 607 bp in length
* 177583 177662: gap of unknown length
* 177663 178111: contig of 449 bp in length
* 178112 178191: gap of unknown length
* 178192 178900: contig of 709 bp in length
* 178901 178980: gap of unknown length
```

## Alignment Scores:

```
Pred. No.: 5.08 Length: 198282
Score: 112.50 Matches: 25
Percent Similarity: 59.26% Conservative: 7
Best Local Similarity: 46.30% Mismatches: 17
Query Match: 20.38% Indels: 5
DB: 2 Gaps: 1
```

US-10-020-441-2 (1-106) x ACC07415 (1-198282)

```
OY 37 ArgAtgProIleAlaGlnThrLeuSerGlyPheAspIleAlaIleValMetLeuAlaGln 56
Db 77724 ACAAGAGCATGACT-----TATGATTTCGACCTGTATTTCAATGTCAG 77768
```

```
OY 57 MetValAsnLeuGlnSerGlyIleArgValIleSerLeuProGlnProSerAspIlePro 76
Db 77769 CGGATTACTTCTGGATGATCAGATTAATGTTATGTTTACCTCAACAGATGATATTCG 77828
```

```
OY 77 ProProGlyThrGlyValPheIleValGlyTyrGlyArgAsp 90
Db 77829 CAGCCCGGTAACACATAGTTTCCACCGCGTGGGAAAAGAT 77870
```

```
RESULT 14
BTTRYPTMR 929 bp mRNA linear NAM 22-MAY-1996
LOCUS B.taurus mRNA for trypsin.
DEFINITION X94982.1 GI:1332446
VERSION X94982.1
KEYWORDS trypsin.
SOURCE cow.
ORGANISM Bos taurus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidea;
Bovidae; Bovinae; Bos.
1 (bases 1 to 929)
Pallaofo, M., Gambacurta, A., Fiorucci, L., Mignogna, G., Barra, D. and
Ascoli, F.
cDNA cloning and primary structure of trypsin from bovine mast
cells, and evidence for the expression of bovine pancreatic trypsin
inhibitor mRNA in the same cells
Eur. J. Biochem. 237 (1), 100-105 (1996)
```

```
JOURNAL MEDLINE 96203914
PUBMED 8620861
REFERENCE 2 (bases 1 to 929)
AUTHORS Ascoli, F.
TITLE Direct Submission
JOURNAL Submitted (11-JAN-1996) F. Ascoli, Department of Experimental
Medicine, and Biochemical Sciences, Tor Vergata University, via di
Tor Vergata 135, 00133 Roma, ITALY
FEATURES
source location/Qualifiers
1..929
/organism="Bos taurus"
/db_xref="taxon:9913"
/cell_type="mast cells"
/tissue_type="liver capsule"
/dev_stage="adult"
<1..714
/EC_number="3.4.21.59"
/codon_start=1
```

```
/product="trypsin"
/protein_id="CAA64438.1"
/db_xref="GI:1332447"
/db_xref="SPRMBL:Q29464"
/translation="GSGMPQVSRVRRYWRHHGGSLIHPQVLTAAHCVPEVHG
PSIRVQLRQHLTYDQQLPISRIITHPNCYSYKRNADLALDLTVNISMVQV
TLPESTFPFGTCWVTGNGVNGNRLLPFPFLKQVFNVSVCDRYHSGLSL
GDNPVIVREBMLCAGDSGRNFCQDSSGSLPVCKVNGTWLQAGVSWGCAKPNRPGI
YTRVSYLDIMIHQVPOGP"
BASE COUNT 162 a 326 c 290 g 151 t
ORIGIN
```

## Alignment Scores:

```
Pred. No.: 0.0573 Length: 929
Score: 108.50 Matches: 22
Percent Similarity: 57.58% Conservative: 16
Best Local Similarity: 33.33% Mismatches: 19
Query Match: 19.66% Indels: 9
DB: 4 Gaps: 1
```

US-10-020-441-2 (1-106) x BTTRYPTMR (1-929)

```
OY 29 ProSerCysMetSerAlaArgGlnArgArgProIleAlaGlnThrLeuSerGlyPheAsp 48
Db 217 CCCACTCTCTACAGCGTTAAG-----AACGGCGCGAC 249
```

```
OY 49 IleAlaIleValMetLeuAlaGlnMetValAsnLeuGlnSerGlyIleArgValIleSer 68
Db 250 ATCGCCCTGCTGAGCTGACAGCTTGATATCTCTCTGACAGCTCAGCCGCTCAC 309
```

```
OY 69 LeuProGlnProSerAspIleProProGlyThrGlyValPheIleValGlyTyrGly 88
Db 310 CTGCCCCCGAGTGGAGACCTTCCCGGGGAGCAGGTGCGTACGCGGCTGGGCG 369
```

```
OY 89 ArgAspAspAsnAspArg 94
Db 370 AACGTGACAAATGAAAG 387
```

```
RESULT 15
M33494 2197 bp DNA linear PRI 04-DEC-1998
LOCUS Homo sapiens trypsin-1 gene, complete cds.
DEFINITION M33494
ACCESSION M33494
VERSION M33494.1 GI:3927804
KEYWORDS
SOURCE Homo sapiens.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 2197)
Vanderslice, P., Ballinger, S.M., Tam, E.K., Goldstein, S.M.,
Caughney, G.H. and Caughney, G.H.
Human mast cell trypsin: multiple cDNAs and genes reveal a
multigene serine protease family
Proc. Natl. Acad. Sci. U.S.A. 87 (10), 3811-3815 (1990)
```

```
JOURNAL MEDLINE 90251647
PUBMED 2187193
REFERENCE 2 (bases 1 to 2197)
AUTHORS Caughney, G.H.
TITLE Direct Submission
JOURNAL Submitted (19-JUN-1990) University of California at San Francisco,
San Francisco, CA, USA
3 (bases 1 to 2197)
Caughney, G.H.
Direct Submission
JOURNAL Submitted (25-NOV-1998) University of California at San Francisco,
San Francisco, CA, USA
REMARK
COMMENT On Nov 25, 1998 this sequence version replaced gi:339976.
Draft entry and computer-readable sequence for [Proc. Natl. Acad.
Sci. U.S.A. (1990) In press] kindly submitted
by P.Vanderslice 02-Apr-1990.
```

```

FEATURES
Source                                Location/Qualifiers
                                         1..2197
                                         /organism="Homo sapiens"
                                         /db_xref="taxon:9606"
                                         /cell_type="mast"
                                         /tissue_type="skin"
                                         /dev_stage="adult"
                                         131..135
CAAT_signal                           /evidence=not_experimental
TATA_signal                           197..201
                                         /evidence=not_experimental
mRNA                                  join(<225..247,458..518,673..844,954..1219,1346..1509,
                                         1595..2072)
                                         /product="trypsinase-I"
                                         <225..247
                                         /number=1
                                         248..457
                                         /note="intron A"
                                         join(458..518,673..844,954..1219,1346..1509,1595..1759)
CDS                                   /note="serine protease"
                                         /note_start=1
                                         /product="trypsinase-I"
                                         /protein_id="AAC83172.1"
                                         /db_xref="GI:339977"
                                         /translation="MNLLILALPYLASPAVAPARPOALQRYGVIGGQAPRSKWP
                                         QVSLRVHGPYMMHFCGSLIHPOWVITLAHCVPDYKDLALNVQLREQHLVYDDLL
                                         PVSRLIVHPQFYTAQIGADIALLELEPVNVSSHVHTVILPPASETFPPMPCVTGM
                                         GVDVDERLPPEPLKQVNPVIMENHIQAKYHGAVTGQDVRIVRDVRCAGNTRSD
                                         SCQGDGSGPLVCKVNGTWLQAGVSWGEGCAQENRPGIYTRVYTLDMIHVYKKR"
                                         458..518
                                         /number=2
                                         519..672
                                         /note="intron B"
                                         673..844
                                         /number=3
                                         845..953
                                         /note="intron C"
                                         954..1219
                                         /number=4
                                         1220..1345
                                         /note="intron D"
                                         1346..1509
                                         /number=5
                                         1510..1594
                                         /note="intron E"
                                         1595..2072
                                         /number=6
                                         789 C      674 G      379 T

BASE COUNT      355 A      789 C      674 G      379 T
ORIGIN

Alignment Scores:
Pred. No.:      0.307      Length:      2197
Score:          105.00      Matches:      32
Percent Similarity: 44.92%      Conservative: 21
Best Local Similarity: 27.12%      Mismatches: 44
Query Match:    19.02%      Indels:      22
DB:              9      Gaps:      3

US-10-020-441-2 (1-106) x M33494 (1-2197)

QY      6      LeuAtgAaNgClYAAaNgInNgIy1LeNshNgInProSeRgIyAlLybVAlAlaPro 25
           |||
Db      893  CTGGCTCCCGGGGCTCTCGAGGGCTGCCCAAGGACCTGAGTGGGATCTCCGCTGCCCA 952
           |||
QY      26  gLy-----TymeTProSeRSeCyMeTSeRa1aNgInaNgAyg----- 38
           |||
Db      953  GGGACGCTCAAGGATCTGGGCGCCCTCAAG--GTGCAACTGGGAGAGCAACCTTACTAC 1011
           |||
QY      39  -----Pro1leAaInThrLeu----- 44
           |||
Db      1012 CAGGACCAAGCTGCTGGCTCGGTCAGACAGAGATGATGTCGACACCCACAGTTCTACACGCCGCCAG 1071
           |||
QY      45  SerGlyPhaAsp1leAla1leValMetLeuAlaGlnMetValAenLeuGlnSerGly1le 64

```

Db	1072	ATCGGACGGCATTCGCCCTCTCGAGACCTCGAGAGACCGCGGTGAACGTCCTCCAGCCAGTC	1131
Qy	65	ArgValIleSerLeuProGlnProSerAspIleProProGlyIThrGlyValPheIle	84
Db	1132	CACACGGTACCCCTGCCTCCCTCGCTCAGAGACCTTCGCCCGGGGATGCCGTCTGGGTC	1191
Qy	85	ValGlyTyrGlyArgAspAspAspAspArgAspProSerArgIysAsnGly	102
Db	1192	ACTGGCTGGGGCGATGTGACATATGATGTGGTCTGGGGACAGTGGAGGTGGG	1245
RESULT 16			
LOCUS	AF099143	2280 bp	DNA linear PRI 09-OCT-2001
DEFINITION	Human sapiens mast cell tryptase beta III gene, complete cds.		
ACCESSION	AF099143		
VERSION	AF099143.1	GI:4336616	
KEYWORDS			
SOURCE	Homo sapiens.		
ORGANISM	Homo sapiens.		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.		
AUTHORS	1 (bases 1 to 2280)		
TITLE	Pallaro, M., Pejo, M.S., Shavesteh, L., Blount, J.L. and Caughey, G.H. Characterization of genes encoding known and novel human mast cell tryptases on chromosome 16p13.3		
JOURNAL	J. Biol. Chem. 274 (6), 3355-3362 (1999)		
MEDLINE	99121069		
PUBMED	9920877		
REFERENCE	2 (bases 1 to 2280)		
AUTHORS	Pallaro, M., Pejo, M.S., Shavesteh, L., Blount, J.L. and Caughey, G.H. Direct Submission		
TITLE	Submitted (19-OCT-1998) CVRI, UCSF, 90 Medical Center Way, San Francisco, CA 94143, USA		
JOURNAL	Location/Qualifiers		
FEATURES	1..2280		
source	/organism="Homo sapiens"		
	/db_xref="taxon:9606"		
	/chromosome="16"		
	/map="16p13.3"		
	join(1..273)..333,487..658,768..1033,1161..1324,1410..1887		
	/product="mast cell tryptase beta III"		
	join(1273)..333,487..658,768..1033,1161..1324,1410..1574)		
	/codon_start=1		
	/product="mast cell tryptase beta III"		
	/protein_id="AAD17859.2"		
	/db_xref="GI:15988540"		
	/translation="MNLILLLALPYLASRAYAPAPRGQALQVGVIGGGEAPRSPWP QVSLKRDYRWMHPCGSLIHQWVLTAKCGPVDKDLALRVQRECHLYQDQL FVSRILVHPQFTAGIDIALLELEEPNVSHTVTLPASERFPPMPGWTGW GDVNDERLPPFPLOKOVPIEMENHICPAKHLGATGDDVRIVDDMLCAANTRD SCGGSGPVLCKVGTWLTAGVNSMGECAPNRPBGIVTRYVYLDWIIHVVKKP"		
BASE COUNT	376 a 817 c 683 g 404 t		
ORIGIN			
Alignment Scores:			
Pred. No.:	0.319	Length:	2280
Score:	105.00	Matches:	32
Percent Similarity:	44.92%	Conservative:	21
Best Local Similarity:	27.12%	Mismatches:	44
Query Match:	19.02%	Indels:	22
DB:	9	Gaps:	3
US-10-020-441-2 (1-106) x AF099143 (1-2280)			
Qy	6	LeuArgAsnGlyAspGlnGlnGlyIleHisAsnGlnProSerGlyValIysValAlaPro	25
Db	707	CTGGCTCCCGGGGTCTCTCTGGGGCGTCCAGGCGCCCTGAGTGGATCTCCCGCTGCCA	766
Qy	26	Gly-----TyrMetProSerGlyMetSerAlaArgGlnArgArg-----	38
Db	767	GGGAGCTCAAGGATCTGGCCCGCCTTAGG-GTGCACCTGGGGAGGACACCTTACTAC	825

Qy 39 -----ProilealaginThreleu----- 44  
 Db 826 CAGACACGAGTCTGCGGATGAGACATGATGTCAGACACAGTTCTACACGCCCCAG 885  
 Qy 45 SerGlyPheAspIleAlaIleValMetLeuAlaGlnMetValAsnLeuGlnSerGlyIle 64  
 Db 886 ATCGAGGCGGACATGCGCTGCTGAGCTGAGGAGCCGCTGACGCTTCAGACGCCACGTC 945  
 Qy 65 ArgValIleSerLeuProGlnProSerAspIleProProGlyThrGlyValPheIle 84  
 Db 946 CACAGGTCACCTGCGCCCTGCTCAGACACTTCCCGGAGATGCCGCTGCTGGTTC 1005  
 Qy 85 ValGlyTYrGlyArgAspAspAspAspArgAspProSerArgIlyAsnGlyIly 102  
 Db 1006 ACTGGCTGGGCGATGTGACATATGTGTGGTCTGGGACAGTCGAGTGGG 1059  
 RESULT 17  
 AF099144 5456 bp DNA linear PRI 10-OCT-2001  
 LOCUS Homo sapiens mast cell tryptase beta 1 gene, complete cds.  
 DEFINITION AF099144  
 ACCESSION AF099144  
 VERSION AF099144.1 GI:4336618  
 KEYWORDS  
 ORGANISM Homo sapiens.  
 SOURCE  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 REFERENCE 1 (bases 1 to 5456)  
 AUTHORS Pallares,M., Pejzo,M.S., Shayesteh,L., Blount,J.L. and Caughey,G.H.  
 TITLE Characterization of genes encoding known and novel human mast cell  
 JOURNAL tryptases on chromosome 16p13.3  
 MEDLINE J. Biol. Chem. 274 (6), 3355-3362 (1999)  
 PUBMED 9920877  
 REFERENCE 2 (bases 1 to 5456)  
 AUTHORS Pallares,M., Pejzo,M.S., Shayesteh,L., Blount,J.L. and Caughey,G.H.  
 TITLE Direct Submission  
 JOURNAL Submitted (19-OCT-1998) CVRI, UCSF, 90 Medical Center Way, San  
 Francisco, CA 94143, USA  
 FEATURES  
 source location/Qualifiers  
 1..5456  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /chromosome="16"  
 /map="16p13.3"  
 repeat\_region  
 1036..1121  
 /rpt\_type=dispersed  
 join(1872..1923,2134..2194,2348..2519,2629..2894,  
 3022..3185,3271..3749)  
 /product="mast cell tryptase beta 1"  
 join(2134..2194,2348..2519,2629..2894,3022..3185,  
 3271..3435)  
 /note="serine protease"  
 /codon\_start=1  
 /product="mast cell tryptase beta 1"  
 /protein\_id="AADI7860.1"  
 /db\_xref="GI:4336619"  
 /translation="MTNLTLLPLTASRAVAPAPGALQVGVIGQGEAPRSKMP  
 QVSLRVGPMYMHFGGSLIHQWVLTAAHCGPVKDLAARVQLREHLYIQQL  
 PVSRIIVHOFYTAQIGADIALLEBPVNVSHVATVLPASSTFPGMVCWTGM  
 GVNDVDERLPFPPLKQVNPVIMENHIIDAKYHLGATGDDVRIVRDMLCAGNRRD  
 SCQDGGGLVCKVNGTWLQAGVSWGECAPNRPGLIYTRVYLDWIHHVPKP"  
 repeat\_region  
 4120..4407  
 /rpt\_family="Aluub"  
 /rpt\_type=dispersed  
 BASE COUNT 951 a 1768 c 1703 g 1025 t 9 others  
 ORIGIN  
 Alignment Scores:  
 Pred. No.: 0.769 Length: 5456  
 Score: 105.00 Matches: 32  
 Percent Similarity: 44.92% Conservative: 21  
 Best Local Similarity: 27.12% Mismatches: 44

Query Match: 19.02% Indels: 22  
 DB: 9 Gaps: 3  
 US-10-020-441-2 (1-106) x AF099144 (1-5456)  
 Qy 6 LeuArgAsnGlyAspGlnGlyIlyIleHisGlnProSerGlyValIlyValAlaPro 25  
 Db 2568 CTGGCTCCCGGATGCTCTCGGGGGGCTCCAGAGGCCCTGAGTGGATCTCCGCTGCCCA 2627  
 Qy 26 Gly-----TyrMetProSerCysMetSerAlaArgIlnArgArg----- 38  
 Db 2628 GGGACCTCAAGATGTGCGCCGCTCAGG-GTGCAACTGCCGGAGACACACTTACTAC 2686  
 Qy 39 -----ProilealaginThreleu----- 44  
 Db 2687 CAGACACGAGTCTGCGGATGAGACATGATCTGTCAGACACAGTTCTACACGCCCCAG 2746  
 Qy 45 SerGlyPheAspIleAlaIleValMetLeuAlaGlnMetValAsnLeuGlnSerGlyIle 64  
 Db 2747 ATCGAGGCGGACATGCGCTGCTGAGCTGAGGAGCCGCTGACGCTTCAGACGCCAGTC 2806  
 Qy 65 ArgValIleSerLeuProGlnProSerAspIleProProGlyThrGlyValPheIle 84  
 Db 2807 CACACGTCACCTGCGCCCTGCTCAGACACTTCCCGGAGATGCCGCTGCTGGTTC 2866  
 Qy 85 ValGlyTYrGlyArgAspAspAspAspArgAspProSerArgIlyAsnGlyIly 102  
 Db 2867 ACTGGCTGGGCGATGTGACATATGTGTGGTCTGGGACAGTCGAGTGGG 2920  
 RESULT 18  
 HS303A1/C  
 LOCUS Human DNA sequence from clone LA16-303A1 on chromosome 16, complete  
 DEFINITION sequence.  
 ACCESSION AL031704  
 VERSION AL031704.24 GI:11878023  
 KEYWORDS HTG.  
 ORGANISM human.  
 SOURCE  
 Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 REFERENCE 1 (bases 1 to 30228)  
 AUTHORS Lloyd,C.  
 TITLE Direct Submission  
 JOURNAL Submitted (16-DEC-2000) Sanger Centre, Hinxton, Cambridgeshire,  
 CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk  
 COMMENT  
 requests: clonerequests@sanger.ac.uk  
 On Dec 17, 2000 this sequence version replaced gi:11230433.  
 During sequence assembly data is compared from overlapping clones.  
 Where differences are found these are annotated as variations  
 together with a note of the overlapping clone name. Note that the  
 variation annotation may not be found in the sequence submission  
 corresponding to the overlapping clone, as we submit sequences with  
 only a small overlap as described above.  
 This sequence has been finished according to sequence map criteria  
 as follows. An attempt is made to resolve all sequencing problems,  
 such as compressions and repeats, but not necessarily within known  
 annotated repeat sequence elements. Where the sequence is  
 ambiguous, there is an annotation using the 'unsure' feature key.  
 The following abbreviations are used to associate primary accession  
 numbers given in the feature table with their source databases:  
 Em., EMBL; Sw., SWISSPROT; Tr., TREMBL; Wp., WORMPEP; Information  
 on the WORMPEP database can be found at  
 http://www.sanger.ac.uk/Projects/C/elegans/wormpep  
 LA16-303A1 is  
 part of a clone contig from the tip of the short arm of chromosome  
 16 spanning 2Mb of p13.3 (Higgs D.R., Flint J., Daniels R., MRC  
 Molecular Haematology Unit, Institute of Molecular Medicine, Oxford  
 (unpublished)), and is from the Los Alamos, flow sorted human  
 chromosome 16 libraries constructed by Norman Doggett  
 (unpublished). VECTOR: scos-1  
 IMPORTANT: This sequence is not the entire insert of clone  
 LA16-303A1. It may be shorter because we sequence overlapping  
 sections only once, except for a 100 base overlap.

FEATURES  
The true left end of clone LA16-303A1 is at 1 in this sequence. The true left end of clone LA16-333E1 is at 30129 in this sequence. The true right end of clone LA16-357D8 is at 1812 in this sequence.

## SOURCE

1. 30228  
Location/Qualifiers

misc\_feature  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="16"  
/clone="LA16-303A1"  
/clone\_1lb="LA16"

misc\_feature  
3136..3197  
/note="Weak data"

misc\_feature  
14153  
/note="Random repeat. Forced join. Gap size estimated to be approximately 200bp by restriction digest data."

misc\_feature  
17718  
/note="Random repeat. Forced join. Gap size estimated to be approximately 400bp by restriction digest data."

misc\_feature  
26789..26882  
/note="Single clone region. In alu. Assembly confirmed by restriction digest data."

misc\_feature  
26904..26982  
/note="Weak data in alu."

BASE COUNT 6183 a 8830 c 9037 g 6178 t  
ORIGIN

## Alignment Scores:

Pred. No.: 4.32 Length: 30228  
Score: 105.00 Matches: 32  
Percent Similarity: 44.92% Conservative: 21  
Best Local Similarity: 27.12% Mismatches: 44  
Query Match: 19.02% Indels: 22  
DB: 9 Gaps: 3

US-10-020-441-2 (1-106) x HS303A1 (1-30228)

OY 6 LeuATgAanglyAspGlnGly11eH1sH1sGlnProSerGlyVallyswAlaAlapro 25

DB 4087 CTGGCTCCCGGGTGTCTCGGGGGCTGCCAGGGCCCTGATGGGATCCTCCGCTGCCA 4028

OY 26 GLY-----TyrMetProSerCyMetSerAlaGlnArg----- 38

DB 4027 GGGAGCTCAGAGATCGCGCCGCTCAGG-GTGCACATCGGGAGACGACCTCTACTAC 3969

OY 39 -----ProLeaLagInThrLeu----- 44

DB 3968 CAGGACCAAGCTCTCGCGGTGAGAGATCATCTGTCACCCACAGTTCTACACCGCCGAG 3909

OY 45 SerGlyPheAspIleAlaIleValMetLeuAlaGlnMetValAsnLeuGlnSerGlyIle 64

DB 3908 ATCGGAGCGGACATGCGCTGCTGAGCTGAGAGACCGGTGACCTCTCCAGCCACGTC 3849

OY 65 ArgValIleSerLeuProGlnProSerArgIleProProGlyThrGlyValPheIle 84

DB 3848 CACACGGTCACCTCCCTCCCTCAGAGACCTTCCCGGGGATGCGGTGCTGCTGTC 3789

OY 85 ValGlyTyrGlyArgAspAspAspArgAspProSerArgLyAsnGlyIle 102

DB 3788 ACTGCTGGGGGATGTGACATATGTGGTCTGGGAGACGTGAGGTGG 3735

RESULT 19

LOCUS HS357D8 36006 bp DNA linear PRI 05-APR-2001

DEFINITION Human DNA sequence from clone LA16-357D8 on chromosome 16, complete

ACCESSION AL031715

VERSION AL031715.31 GI:13560066

KEYWORDS HTG.

SOURCE human.

ORGANISM Homo sapiens

REFERENCE 1 (bases 1 to 36006)

AUTHORS Lloyd, C.  
TITLE Direct Submission  
JOURNAL Submitted (05-APR-2001) Sanger Centre, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk

## COMMENT

request: clonerequest@sanger.ac.uk  
On Apr 6, 2001 this sequence version replaced gi:11322236.  
During sequence assembly data is compared from overlapping clones. Where differences are found these are annotated as variations together with a note of the overlapping clone name. Note that the variation annotation may not be found in the sequence submission corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above.  
This sequence was finished as follows unless otherwise noted: all regions were either double-stranded or sequenced with an alternate chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one plasmid subclone or more than one M13 subclone; and the assembly was confirmed by restriction digest. The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: Em: EMBL; Sw: SWISSPROT; Tr: TREMBL; Wp: WORMPEP; information on the WORMPEP database can be found at [http://www.sanger.ac.uk/Projects/C\\_elegans/wormpep](http://www.sanger.ac.uk/Projects/C_elegans/wormpep)  
LA16-357D8 is part of a clone contig from the tip of the short arm of chromosome 16 spanning 2Mb of p13.3 (Higgs D.R., Flint J., Daniels R., MRC Molecular Haematology Unit, Institute of Molecular Medicine, Oxford (unpublished)), and is from the Los Alamos flow sorted human chromosome 16 libraries constructed by Norman Doggett (unpublished). VECTOR: sCos-1  
IMPORTANT: This sequence is not the entire insert of clone LA16-357D8 it may be shorter because we sequence overlapping sections only once, except for a 100 base overlap.  
The true left end of clone LA16-357D8 is at 1 in this sequence. The true left end of clone LA16-303A1 is at 3597 in this sequence. The true right end of clone LA16-320B12 is at 12990 in this sequence.

## FEATURES

## SOURCE

1. 36006  
Location/Qualifiers

misc\_feature  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="16"  
/clone="LA16-357D8"  
/clone\_1lb="LA16"

repeat\_region  
1387..1465

repeat\_region  
1860..1951

repeat\_region  
1866..1957

repeat\_region  
2370..2471

repeat\_region  
4031..8078

repeat\_region  
5262..5321

repeat\_region  
7526..7932

repeat\_region  
9087..9188

repeat\_region  
10107..10616

repeat\_region  
12832..13561

repeat\_region  
13632..13891

repeat\_region  
13633..13845

repeat\_region  
14342..14465

repeat\_region  
14342..14465

repeat\_region  
14342..14465

```
misc_feature /note="4 copies 31 mer 82% conserved"
15075..16589
/note="Cpg island"
evidence=not_experimental
repeat_region 18702..18797
/note="LIMC4 repeat: matches 6647..6744 of consensus"
19888..19944
/note="3 copies 19 mer 87% conserved"
19889..19982
/note="47 copies 2 mer cc 63% conserved"
21793..21894
/note="MLTIF repeat: matches 304..411 of consensus"
22632..22922
/note="ALuB repeat: matches 1..293 of consensus"
23359..23510
/note="8 copies 19 mer 75% conserved"
23809..23850
/note="21 copies 2 mer gg 76% conserved"
25765..26726
/note="LIM3 repeat: matches 146..297 of consensus"
26379
/note="Random repeat. Forced join. Gap size estimated to
be approximately 400bp by restriction digest data."
26442..26566
/note="Weakly double stranded."
27228..27648
/note="Cpg island"
evidence=not_experimental
repeat_region 27488..27942
/note="LIMC repeat: matches 603..1068 of consensus"
27943..28236
/note="ALuSg repeat: matches 1..296 of consensus"
28237..28280
/note="LIMC repeat: matches 1068..1110 of consensus"
28285..28416
/note="ALuSg/x repeat: matches 4..135 of consensus"
28417..28720
/note="ALu repeat: matches 1..303 of consensus"
28721..29007
/note="LIMC repeat: matches 1946..2242 of consensus"
29019..29520
/note="LIMC2 repeat: matches 5828..6328 of consensus"
29521..29832
/note="ALu repeat: matches 1..309 of consensus"
29833..30369
/note="LIMC2 repeat: matches 5280..5828 of consensus"
30343..30384
/note="LIM1 repeat: matches 851..892 of consensus"
30415..30722
/note="ALuSg repeat: matches 1..308 of consensus"
30767..30922
/note="LIMC2 repeat: matches 5156..5312 of consensus"
30928..31049
/note="ALuB/FRAM repeat: matches 169..290 of consensus"
31211..31321
/note="LIMC2 repeat: matches 6048..6164 of consensus"
31759..31873
/note="5 copies 23 mer 68% conserved"
31782..31876
/note="5 copies 19 mer 71% conserved"
31783..31878
/note="48 copies 2 mer gg 67% conserved"
32512
32514
32554..32619
/note="LIM repeat: matches 5408..5473 of consensus"
32912..33185
/note="137 copies 2 mer tg 71% conserved"
32923..33194
/note="8 copies 34 mer 70% conserved"
33121..33262
/note="2 copies 71 mer 83% conserved"
33193..33406
repeat_region
```

```
repeat_region /note="107 copies 2 mer tg 64% conserved"
33209..33412
/note="6 copies 34 mer 68% conserved"
repeat_region 33477..33599
/note="LIMC/D repeat: matches 5276..5401 of consensus"
33647..33828
/note="LIMC3 repeat: matches 6834..7013 of consensus"
34028..34177
/note="LIMC3 repeat: matches 7304..7460 of consensus"
34337..34432
/note="48 copies 2 mer ct 61% conserved"
34351..34419
/note="3 copies 23 mer 81% conserved"
34442..34557
/note="ALuB repeat: matches 186..312 of consensus"
34518..34653
/note="ALuB repeat: matches 158..302 of consensus"
34668..34966
/note="ALuSg repeat: matches 1..299 of consensus"
34970..35019
/note="LIMD3 repeat: matches 6939..6993 of consensus"
35020..35311
/note="ALuSg repeat: matches 17..313 of consensus"
35312..35402
/note="LIMD3 repeat: matches 6993..7064 of consensus"
35396..35716
/note="match: GSS: Em:AQ349283"
35406..35781
/note="match: GSS: Em:AQ339836"
35722..35905
/note="LIMC3 repeat: matches 7292..7485 of consensus"
BASE COUNT 6824 a 11086 c 11379 g 6717 t
ORIGIN
Alignment Scores:
Pred. No.: 5.16 Length: 36006
Score: 105.00 Matches: 32
Percent Similarity: 44.92% Conservative: 21
Best Local Similarity: 27.12% Mismatches: 44
Query Match: 19.02% Indels: 22
DB: 9 Gaps: 3
US-10-020-441-2 (1-106) x HS357D8 (1-36006)
QY 6 LeuArgAsnGlyAAspGlnGlyIleHisGlnProSerGlyValIysValAlaPro 25
DB 24473 CTGGCTCCCGGGGCTCTCTCGGGGGCTCGCCAGGCGCCCTGAGTGGATCTCCGGCTGCCA 24414
QY 26 Gly-----TyrMetProSerCysMetSerAlaArgGlnArg- 38
DB 24413 GGGAGGTCAAGATCTGGCGCGCTCCAGG-GTGCAACTCGGGAGACAGCACTTACTAC 24355
QY 39 -----ProIleAlaGlnThrLeu----- 44
DB 24354 CAGAGCCAGCTGCTCGCTCGGTCAGAGATCATGTGCACCCACAGTTCTACACCGCCAG 24295
QY 45 SerGlyPheAspIleAlaIleValMetLeuAlaGlnMetValAsnLeuInsGlyIle 64
DB 24294 ATCGGACGGACATCGCTCGCTGAGCTGAGAGACCGGTAAAGCTCTCCAGCCACGTC 24235
QY 65 ArgValIleSerLeuProGlnProSerAspIleProProGlyThrGlyValPheIle 84
DB 24234 CACACGCTACCTGCGCCCTCGCTCAGAGACCTTCCCGGGGATGCGGTGCTGGTC 24175
QY 85 ValGlyTyrGlyArgAspAsnAspAspArgAspProSerArgIysAsnGlyIle 102
DB 24174 ACTGCTGGGGGATGTGACATGATGTGGTGTGGGAGACAGTGGAGTGGG 24121
RESULT 20
AC120498 AC120498 195680 bp DNA linear PRI 08-JUL-2002
LOCUS Homo sapiens chromosome 16 clone RP11-616M22, complete sequence.
ACCESSION AC120498 AC084177
```



source 206827. .250642  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="cosmid 333e1"  
246926. .265786  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="cosmid 358b7"  
<1. .1485  
/note="HS313F9"  
repeat\_region  
/note="AluUo; RepeatMasker predicted  
23/10/2000"  
/rpt\_family="SINE/Alu"  
complement(280. .664)  
repeat\_region  
/note="MLTIC; RepeatMasker predicted  
23/10/2000"  
/rpt\_family="LTR/MaLR"  
misc\_feature  
837. .1937  
/note="Cpg island; not associated with any known gene"  
evidence=not\_experimental  
1386. .14397  
/note="HS305F3"  
misc\_feature  
2205. .2565  
/note="MST1; RepeatMasker predicted  
23/10/2000"  
/rpt\_family="LTR/MaLR"  
repeat\_region  
2567. .2598  
/note="(TTTG)n; RepeatMasker predicted  
23/10/2000"  
/rpt\_family="Simple repeat"  
complement(2616. .2906)  
repeat\_region  
/note="AluSg; RepeatMasker predicted  
23/10/2000"  
/rpt\_family="LTR/ERV.L"  
repeat\_region  
3241. .3337  
/note="(TTG)n; RepeatMasker predicted  
23/10/2000"  
/rpt\_family="Simple repeat"  
3350. .3515  
/note="LTR32; RepeatMasker predicted  
23/10/2000"  
/rpt\_family="LTR/ERV.L"  
repeat\_region  
4698. .4748  
/note="MIR; RepeatMasker predicted  
23/10/2000"  
/rpt\_family="SINE/MIR"  
complement(7543. .7687)  
repeat\_region  
/note="MERSA; RepeatMasker predicted  
23/10/2000"  
/rpt\_family="DNA/MER1\_type"  
8207. .8232  
/note="(CCCG)n; RepeatMasker predicted  
23/10/2000"  
/rpt\_family="SINE/Alu"  
11335. .11664  
/note="CT-rich; RepeatMasker predicted  
23/10/2000"  
/rpt\_family="Low complexity"  
complement(12201. .12427)  
repeat\_region  
/note="LIMed; RepeatMasker predicted  
23/10/2000"  
/rpt\_family="LINE/L1"  
complement(12994. .1386)  
repeat\_region  
/note="AluUb; RepeatMasker predicted

23/10/2000"  
/rpt\_family="SINE/Alu"  
13787. .15957  
/note="Cpg island; SSTR5"  
evidence=not\_experimental  
14298. .47299  
/note="HS349E11"  
gene  
14397. .15644  
/gene="SSTR5"  
14397. .15644  
/gene="SSTR5"  
/note="Similar to Unigene cluster Hs241375"  
14432. .15526  
CDS  
/note="PubMed:7908405"  
/gene="SSTR5"  
/product="somatostatin receptor type 5"  
/protein\_id="AAK61266.1"  
/db\_xref="GI:1436736"  
/db\_xref="MIM:182455"  
/db\_xref="SWISS-PROT:P35346"  
/translation="MEPLFPASTPSMNASPGASGGDNRRTLVGPASAGARAVLP  
VLYLVCAGIGNTIVYVLRRAKKVTNITILNLAVDVLYMGLPELTANAA  
SFMPFGVLCILVNTLDGVNPFVSFCITWSVRYLAVVHPLSSARRPRVAKLS  
AAAVVLSLMSLPLVPAVDVGGTCNAMPVGLGAVFTITAVIGFPAFLVTC  
LCYLLIVKVRPAAGVRGCVRRSERKVRVAVLVAFACWLPFTVNLAVAL  
POEPASAGLYFFVILSYANSCANPVLYGFLSDNFRSOFKVLCLKRSGAKADATE  
LRPDRIRGOOATPPAHRANGLMOTSKL"  
complement(18028. .18082)  
repeat\_region  
/note="L2; RepeatMasker predicted  
23/10/2000"  
/rpt\_family="LINE/L2"  
19693. .19971  
/note="AluUo; RepeatMasker predicted  
23/10/2000"  
/rpt\_family="SINE/Alu"  
complement(21110. .21550)  
repeat\_region  
/note="L1MB3; RepeatMasker predicted  
23/10/2000"  
/rpt\_family="LINE/L1"  
complement(21580. .21658)  
repeat\_region  
/note="MLT1A1; RepeatMasker predicted  
23/10/2000"  
/rpt\_family="LTR/MaLR"  
complement(21772. .22038)  
repeat\_region  
/note="AluSg; RepeatMasker predicted  
23/10/2000"  
/rpt\_family="SINE/Alu"  
complement(22041. .22115)  
repeat\_region  
/note="MLT1A1; RepeatMasker predicted  
23/10/2000"  
/rpt\_family="LTR/MaLR"  
complement(22117. .22164)  
repeat\_region  
/note="L1MB3; RepeatMasker predicted  
23/10/2000"  
/rpt\_family="LINE/L1"  
complement(22218. .22499)  
repeat\_region  
/note="AluSx; RepeatMasker predicted  
23/10/2000"  
/rpt\_family="SINE/Alu"  
23130. .23422  
/note="AluUo; RepeatMasker predicted  
23/10/2000"

Alignment Scores:  
Pred. No.: 38.7  
Score: 105.00  
Percent Similarity: 44.92%  
Best Local Similarity: 27.12%  
Length: 265786  
Matches: 32  
Conservative: 21  
Mismatches: 44



Query Match: 19.02% Indels: 22  
 DB: 9 Gaps: 3  
 US-10-020-441-2 (1-106) x AB006466 (1-265786)  
 QY 6 LeuAArgAsnGlyAAspGlnGlnGlyIleHisHisGlnProSerGlyValIysValAlaPro 25  
 DB 180686 CTGGGCTCCCGGGGCTCTCGGGGCTGCCAGAGCCCTGCTGGATCCCGCTGCCCA 180627  
 QY 26 Gly-----TyrMetProSerGlyMetSerAlaArgGlnArg----- 38  
 DB 180626 GGGAGCTCAAGATCTGGCCGCTTCAAG-GTCAACTGCGGAGAGACCTCTACTAC 180568  
 QY 39 -----ProIleAlaGlnThrLeu----- 44  
 DB 180567 CAGAGACCACTGCTGCGGCTGACAGATCATGTCACCCACAGTTCTACACCGCCAG 180508  
 QY 45 SerGlyPheAspIleAlaIleValMetLeuAlaGlnMetValAsnLeuGlnSerGlyIle 64  
 DB 180507 ATCGGAGCGGACATCGCCCTGCTGAGAGACCGGTTGAACGCTCTCACCCACAGTC 180448  
 QY 65 ArgValIleSerLeuProGlnProSerAspIleProProGlyIleThrGlyValPheIle 84  
 DB 180447 CACACGGTACCTGCTGCCCTGCTCAGAGACCTTCCCTCCGAGATGCCGTCGAGTC 180388  
 QY 85 ValGlyTyrGlyArgAspAsnAspArgAspProSerArgIysAsnGlyIle 102  
 DB 180387 ACTGGCTGGGCGCATGTGACATGATGTGGTCTGGGAGACAGTGAAGTGGG 180334  
 RESULT 22  
 OARI8223 1002 bp mRNA linear MAM 09-JUN-2000  
 LOCUS Ovis aries mRNA for trypsin-1.  
 DEFINITION Y18223  
 ACCESSION Y18223.1 GI:4753834  
 VERSION Y18223.1 GI:4753834  
 KEYWORDS trypsin.  
 SOURCE sheep.  
 ORGANISM Ovis aries  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;  
 Bovidae; Caprinae; Ovis.  
 1 (bases 1 to 1002)  
 Pemberton, A.D., McAlene, S.M., Huntley, J.F., Collie, D.D.,  
 Scudamore, C.L., McEuen, A.R., Wallis, A.F., and Miller, H.R.,  
 cDNA sequence of two sheep mast cell trypsinases and the differential  
 expression of trypsinase and sheep mast cell proteinase-1 in lung,  
 dermis and gastrointestinal tract  
 Clin. Exp. Allergy 30 (6), 818-832 (2000)  
 JOURNAL MEDLINE 20308142  
 PUBMED 10848900  
 REFERENCE 2 (bases 1 to 1002)  
 AUTHORS McAlene, S.M.  
 TITLE Direct Submission  
 JOURNAL Submitted (13-OCT-1998) S.M. McAlene, University of Edinburgh,  
 R(D)SV, Veterinary Clinical Studies, Easter Bush Veterinary Centre,  
 Roslin, Midlothian, EH25 9RG, UK  
 FEATURES  
 SOURCE 1.1002  
 /organism="Ovis aries"  
 /db\_xref="taxon:9940"  
 /cell\_type="mast cell"  
 /tissue\_type="bone marrow"  
 /EC\_number="3.4.21.59"  
 /note="trypsin-1"  
 /codon\_start=1  
 /product="trypsinase"  
 /protein\_id="CAB41988.1"  
 /db\_xref="GI:4753835"  
 /db\_xref="SPTREMBL:O9XSM1"  
 /translation="MLHLALALLSLVSAAPGQALQSGITGKEAPGSRWQY  
 SLRVDOYWRHQCGSLIPQWILPAICIGELPEQSPDFRVLAEQHLHYODRLPI  
 SRVLPHPYVENVGADIALQLLEPVSISCHVRVTLPPASETPPPSQCMVTGWN

mat\_peptide  
 polyA\_signal  
 polyA\_site  
 BASE COUNT 173 a 357 c 312 g 159 t 1 others  
 ORIGIN  
 Alignment Scores:  
 Pred. No.: 0.248 Length: 1002  
 Score: 102.50 Matches: 31  
 Percent Similarity: 43.90% Conservative: 23  
 Best Local Similarity: 25.20% Mismatches: 36  
 Query Match: 18.57% Indels: 3  
 DB: 4 Gaps: 4  
 US-10-020-441-2 (1-106) x OARI8223 (1-1002)  
 QY 4 LeuThrLeuArgAsnGlyAAspGlnGlnGlyIleHisHisGlnProSerGlyValIysVal 23  
 DB 155 GTGAGCTGAGAGTACAGAGACCACTGAG--AGGACACAGTGGCGGCTCCCTGATC 211  
 QY 24 AlaProGlyTyrMet-----Proser 30  
 DB 212 CACCCCACTGGGTGTGACCCGACCCACTGATCGAGACCGGAACCTCCAGAGCCCTTA 271  
 QY 31 CysMetSerAlaArgGlnArg-----ProIle 40  
 DB 272 GACTTCAGGCTGAGGTGGGAGAGACACCTGACTACAGACCGGCTGCTGCCTATC 331  
 QY 41 AlaGlnThrLeu-----SerGlyPheAspIleAlaIle 51  
 DB 332 AGCAGGGTATCCCCACCCCACTACTACATGTTGAATGGGCGGACATCGCCCTG 391  
 QY 52 ValMetLeuAlaGlnMetValAsnLeuGlnSerGlyIleArgValIleSerLeuProGln 71  
 DB 392 CTGACAGCTGAGAGAGCCCTGAGCATCTCTCCATGTCGCCGCGTACCTGCCCCCT 451  
 QY 72 ProSerAspIleProProGlyIleThrGlyValPheIleValGlyTyrGlyArgAsp 91  
 DB 452 GGATGGAGACCTTCCCTCCAGGTCACAGCTGCTGGGAGCGGCTGAGCAACGTGAC 511  
 QY 92 AsnAspArg 94  
 DB 512 AATGGAAG 520  
 RESULT 23  
 AF361940 3516 bp DNA linear ROD 07-MAY-2001  
 LOCUS Mus musculus mast cell protease 10 (Mcp10) gene, complete cds.  
 DEFINITION AF361940  
 ACCESSION AF361940.1 GI:13959031  
 VERSION AF361940.1 GI:13959031  
 KEYWORDS  
 SOURCE Mus musculus.  
 ORGANISM Mus musculus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 1 (bases 1 to 3516)  
 Hunt, J.E., McNeill, P., Austen, F., Gurish, M., and Stevens, R.L.  
 Molecular cloning of the cDNA and gene encoding mouse mast cell  
 protease 10  
 JOURNAL Unpublished  
 REFERENCE 2 (bases 1 to 3516)  
 AUTHORS Hunt, J.E., McNeill, P., Austen, F., Gurish, M., and Stevens, R.L.  
 TITLE Direct Submission  
 JOURNAL Submitted (15-MAR-2001) Pathology, University of New South Wales,  
 High St, Sydney, NSW 2052, Australia  
 FEATURES  
 SOURCE 1.3516  
 /organism="Mus musculus"



TITLE	Larsson, L.G., Nilsson, K. and Hellman, L.
JOURNAL	Expression of a mast cell tryptase in the human monocytic cell
MEDLINE	lines U-937 and Mono Mac 6
PUBMED	Scand. J. Immunol. 38 (4), 359-367 (1993)
	94023807
	8210998

```

sig_peptide      6. .68
                  /gene="MMP-6"
mac_peptide     99. .833
                  /gene="MMP-6"
                  /product="serine protease"
polyA_signal    1063. .1068
                  /gene="MMP-6"
BASE COUNT      217 a      330 c      289 g      254 t
ORIGIN

```

ACCESSION	AR080464	GI:10007199
VERSION	AR080464.1	
KEYWORDS	Unknown.	
SOURCE	Unknown.	
ORGANISM	Unclassified.	
REFERENCE	1 (bases 1 to 1108)	
AUTHORS	Stevens,R.L.	
TITLE	Mast cell protease that cleaves fibrinogen	
JOURNAL	Patent: US 5968782-A 20 19-OCT-1999;	
FEATURES	Location/Qualifiers	
source	1..1108	/organism="unknown"
BASE COUNT	221 a 333 c 297 g 257 t	
ORIGIN		
Alignment Scores:		
Pred. No.:	0.308	Length: 1108
Score:	102.00	Matches: 21
Percent Similarity:	62.50%	Conservative: 9
Best Local Similarity:	43.75%	Mismatches: 18
Query Match:	18.48%	Indels: 0
DB:	6	Gaps: 0
US-10-020-441-2 (1-106) x AR080464 (1-1108)		
Oy	46 GlyPheaspIleAlaIleValMetIeuAlaGlnMetValAsnLeuGlnSerGlyIleArg 65	
Dd	391 GGGCGAGCGTTGCCCTGTGCAGCTTCCCTGTGAATGTCTCCACCATTATCCAC 450	
Oy	66 ValIleSerIeuProGlnProSerAspIleProProGlyThrGlyValPheIleVal 85	
Dd	451 CCCATATCCTGCCCTGCCCTCGAGACTTCCTCCCTGGAGCATGCTGGTGACA 510	
Oy	86 GlyTYrGlyArgaspAspAsnasp 93	
Dd	511 GGCTGGGGCGACATTGATAATGAC 534	
RESULT 28		
LOCUS	MUSPROTB 1108 bp mRNA linear ROD 27-APR-1993	
DEFINITION	Mouse protease-6 gene, complete cds.	
ACCESSION	M57626	
VERSION	M57626.1 GI:200508	
KEYWORDS	protease 6.	
SOURCE	Mouse (strain DBA/2) mast cell line K1SV-MC, CDNA to mRNA.	
ORGANISM	Mus musculus	
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.	
AUTHORS	1 (bases 1 to 1108)	
TITLE	Reynolds,D.S., Gurley,D.S., Austen,K.F. and Serafin,W.E.	
JOURNAL	Cloning of the cDNA and gene of mouse mast cell protease-6.	
FEATURES	Transcription by progenitor mast cells and mast cells of the connective tissue subclass	
source	J. Biol. Chem. 266 (6), 3847-3853 (1991)	
MEDLINE	91139682	
PUBMED	1995638	
FEATURES		
source	1..1108	Location/Qualifiers
	/organism="Mus musculus"	
	/strain="DBA/2"	
	/db_xref="taxon:10090"	
	/cell_line="K1SV-MC"	
	/cell_type="mast"	
	1..1108	
gene	/gene="protease-6"	
mRNA	1..1108	
	/gene="protease-6"	
CDs	34..864	
	/gene="protease-6"	
	/codon_start=1	
	/product="protease-6"	
	/protein_id="AAAJ9988.1"	

/db_xref="GI:200509"		
/translation="MLKRGLLLALSLASLVYSAPPANQRVIGVGEHSESKMP		
MOVLSRFKLNTWIFRCGSLIRHPQLFPAACVGHPIKSPLFRVOLREOVLYYGDOL		
LSLNRIYVPHPTYTEGGADVALLETTPAVNVSTHIHIPISLPSAEFPFGSCMTVG		
WDIDNDEPLPPYPFKOVPIVENSLCDRKHTTGLTGDPDFTIHDMCLCAENTRR		
DSCOGDSGGPIVCVKVTQWTLOGVVSWEGCAQPNKPGIYTRVITYLDWIHRIVPEHS		
"		
sig_peptide	34..96	/gene="protease-6"
mat_peptide	127..861	/gene="protease-6"
	/product="protease-6"	
BASE COUNT	221 a 333 c 297 g 257 t	
ORIGIN		
Alignment Scores:		
Pred. No.:	0.308	Length: 1108
Score:	102.00	Matches: 21
Percent Similarity:	62.50%	Conservative: 9
Best Local Similarity:	43.75%	Mismatches: 18
Query Match:	18.48%	Indels: 0
DB:	10	Gaps: 0
US-10-020-441-2 (1-106) x MUSPROTB (1-1108)		
Oy	46 GlyPheaspIleAlaIleValMetIeuAlaGlnMetValAsnLeuGlnSerGlyIleArg 65	
Dd	391 GGGCGAGCGTTGCCCTGTGCAGCTTCCCTGTGAATGTCTCCACCATTATCCAC 450	
Oy	66 ValIleSerIeuProGlnProSerAspIleProProGlyThrGlyValPheIleVal 85	
Dd	451 CCCATATCCTGCCCTGCCCTCGAGACTTCCTCCCTGGAGCATGCTGGTGACA 510	
Oy	86 GlyTYrGlyArgaspAspAsnasp 93	
Dd	511 GGCTGGGGCGACATTGATAATGAC 534	
RESULT 29		
LOCUS	ARI12352 2218 bp DNA linear PAT 16-MAY-2001	
DEFINITION	Sequence 5 from patent US 6130052.	
ACCESSION	ARI12352	
VERSION	ARI12352.1 GI:14092252	
KEYWORDS	Unknown.	
SOURCE	Unknown.	
ORGANISM	Unclassified.	
REFERENCE	1 (bases 1 to 2218)	
AUTHORS	Van Baren,N., Coule,P.G., Desmet,C., Lucas,S. and Boon-Falleur,T.	
TITLE	Leukemia associated genes	
JOURNAL	Patent: US 6130052-A 5 10-OCT-2000;	
FEATURES	Location/Qualifiers	
source	1..2218	/organism="unknown"
BASE COUNT	392 a 793 c 659 g 374 t	
ORIGIN		
Alignment Scores:		
Pred. No.:	0.621	Length: 2218
Score:	102.00	Matches: 35
Percent Similarity:	48.57%	Conservative: 16
Best Local Similarity:	33.33%	Mismatches: 35
Query Match:	18.48%	Indels: 19
DB:	6	Gaps: 4
US-10-020-441-2 (1-106) x ARI12352 (1-2218)		
Oy	9 GlyaPngIngInglyIleHisGlnProSerGly-----Vallys 22	
Dd	223 GGGCGACTGCGTGGAACCGGACATCAAGATCTGCCCGCTCAGGGTGCACTGGCGGA 282	
Oy	23 ValaIaPProGlyTyMetProSerCyMetSerAlaArgGlnArgProIleAlaGln 42	

Db 283 GCAGACCT--CTACTACAGACGAGCTGCTCCGCTGACAGATCATCTGACACC 339  
Qy 43 Threleu-----SerglyPheaplleaileValMetLeuAglmEva 58  
Db 340 ACGATTCTACATCATCCAGACCGGGCGGACATCGCCCTGCTGAGACTGAGAGACCCCT 399  
Qy 58 1AenleuInserGlyIleArgValIleSerleuProGlnProSerAspIleProProPr 78  
Db 400 GAACATCTCCACCATCCACACAGCTGACGCTGCCCTGCTCGAGACCTTCCCCC 459  
Qy 78 OGlyThrglyValPheIleValGlyTYrglyArgAspAspAspAspAspProSerAr 98  
Db 460 GGGATGCCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 502  
Qy 98 glysAenglygly 102  
Db 503 ----AATGTTGGG 511  
RESULT 30  
AR163667 2218 bp DNA linear PAT 17-OCT-2001  
LOCUS AR163667  
DEFINITION Sequence 5 from patent US 6271019.  
ACCESSION AR163667  
VERSION AR163667.1 GI:16234339  
KEYWORDS  
SOURCE Unknown.  
ORGANISM Unknown.  
REFERENCE 1 (bases 1 to 2218)  
AUTHORS Van Baren,N., Coullie,P.G., Desmet,C., Lucas,S. and Boon-Falleur,T.  
TITLE Leukemia associated genes  
JOURNAL Patent: US 6271019-A 5 07-AUG-2001;  
FEATURES  
source 1..2218  
location/Qualifiers  
BASE COUNT 392 a 793 c 659 g 374 t  
ORIGIN  
Alignment Scores:  
Pred. No.: 0.621 Length: 2218  
Score: 102.00 Matches: 35  
Percent Similarity: 48.57% Conservative: 16  
Best Local Similarity: 33.33% Mismatches: 35  
Query Match: 18.48% Indels: 19  
Gaps: 4  
US-10-020-441-2 (1-106) x AR163667 (1-2218)  
Qy 9 G1yAepGInGInGlyIleHIShISGInProSerGly-----Vallys 22  
Db 223 GGGCGACGCGTGGGAACCGGACATCAAGATCTGCGCCGCTCAGAGGTGCACTGCGGGA 282  
Qy 23 VALAlaProGlyTYrMetProSerCyMetSerAlaArgGlnArgProIleAglIn 42  
Db 283 GCAGACCT--CTACTACAGACGAGCTGCTGCCGCTGACAGATCATCTGACACC 339  
Qy 43 Threleu-----SerglyPheaplleaileValMetLeuAglmEva 58  
Db 340 ACGATTCTACATCATCCAGACCGGGCGGACATCGCCCTGCTGAGACTGAGAGACCCCT 399  
Qy 58 1AenleuInserGlyIleArgValIleSerleuProGlnProSerAspIleProProPr 78  
Db 400 GAACATCTCCACCATCCACACAGCTGACGCTGCCCTGCTCGAGACCTTCCCCC 459  
Qy 78 OGlyThrglyValPheIleValGlyTYrglyArgAspAspAspAspAspProSerAr 98  
Db 460 GGGATGCCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 502  
Qy 98 glysAenglygly 102  
Db 503 ----AATGTTGGG 511  
RESULT 31

AR112351  
LOCUS AR112351 2259 bp DNA linear PAT 16-MAY-2001  
DEFINITION Sequence 3 from patent US 6130052.  
ACCESSION AR112351  
VERSION AR112351.1 GI:14092251  
KEYWORDS  
SOURCE Unknown.  
ORGANISM Unknown.  
REFERENCE 1 (bases 1 to 2259)  
AUTHORS Van Baren,N., Coullie,P.G., Desmet,C., Lucas,S. and Boon-Falleur,T.  
TITLE Leukemia associated genes  
JOURNAL Patent: US 6130052-A 3 10-OCT-2000;  
FEATURES  
source 1..2259  
location/Qualifiers  
BASE COUNT 391 a 835 c 642 g 391 t  
ORIGIN  
Alignment Scores:  
Pred. No.: 0.632 Length: 2259  
Score: 102.00 Matches: 35  
Percent Similarity: 48.57% Conservative: 16  
Best Local Similarity: 33.33% Mismatches: 35  
Query Match: 18.48% Indels: 19  
Gaps: 4  
US-10-020-441-2 (1-106) x AR112351 (1-2259)  
Qy 9 G1yAepGInGInGlyIleHIShISGInProSerGly-----Vallys 22  
Db 223 GGGCGACGCGTGGGAACCGGACATCAAGATCTGCGCCGCTCAGAGGTGCACTGCGGGA 282  
Qy 23 VALAlaProGlyTYrMetProSerCyMetSerAlaArgGlnArgProIleAglIn 42  
Db 283 GCAGACCT--CTACTACAGACGAGCTGCTGCCGCTGACAGATCATCTGACACC 339  
Qy 43 Threleu-----SerglyPheaplleaileValMetLeuAglmEva 58  
Db 340 ACGATTCTACATCATCCAGACCGGGCGGACATCGCCCTGCTGAGACTGAGAGACCCCT 399  
Qy 58 1AenleuInserGlyIleArgValIleSerleuProGlnProSerAspIleProProPr 78  
Db 400 GAACATCTCCACCATCCACACAGCTGACGCTGCCCTGCTCGAGACCTTCCCCC 459  
Qy 78 OGlyThrglyValPheIleValGlyTYrglyArgAspAspAspAspAspProSerAr 98  
Db 460 GGGATGCCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 502  
Qy 98 glysAenglygly 102  
Db 503 ----AATGTTGGG 511  
RESULT 32  
AR163666 2259 bp DNA linear PAT 17-OCT-2001  
LOCUS AR163666  
DEFINITION Sequence 3 from patent US 6271019.  
ACCESSION AR163666  
VERSION AR163666.1 GI:16234338  
KEYWORDS  
SOURCE Unknown.  
ORGANISM Unknown.  
REFERENCE 1 (bases 1 to 2259)  
AUTHORS Van Baren,N., Coullie,P.G., Desmet,C., Lucas,S. and Boon-Falleur,T.  
TITLE Leukemia associated genes  
JOURNAL Patent: US 6271019-A 3 07-AUG-2001;  
FEATURES  
source 1..2259  
location/Qualifiers  
BASE COUNT 391 a 835 c 642 g 391 t  
ORIGIN

Alignment Scores:

Pred. No.: 0.632 Length: 2259  
Score: 102.00 Matches: 35  
Percent Similarity: 48.57% Conservative: 16  
Best Local Similarity: 33.33% Mismatches: 35  
Query Match: 18.48% Indels: 19  
DB: 6 Gaps: 4

US-10-020-441-2 (1-106) x AR163666 (1-2259)

QY 9 GYABGNGINGLYIYLEHISGINPROSEGLY-----Vallys 22  
Db 223 GGGCGACCTCGGAGACCGGACATCAAGATCTGGCCGCTCAGAGGTCAACCTGGGGA 282  
QY 23 Val1aProGlyTYrMeProSerCysMetSerAlaArgGlnArgProIlealagin 42  
Db 283 GGAGACCTCTACTACACGAGACCAAGCTGCTCCGCGTACAGAGATATATCTGCACCC 339  
QY 43 Thleu-----SerGlyPheAspIleAlaIleValMetLeuAlaGlnMeVa 58  
Db 340 ACAGTTCATCATCATCCAGACCGGGGCGGACATCCGCTCTGAGAGCTGAGAGCCCGT 399  
QY 58 1AenLeuGlnSerGlyIleArgValIleSerLeuProGlnProSerAspIleProPro 78  
Db 400 GAACATCTCCACCCACATCCACACGCTCAGCGTCCGCTCCGCTCGAGACCTTCCGCC 459  
QY 78 oGlyThrGlyValIleValIleGlyTYrGlyArgAspAspAspAspAspAsp 98  
Db 460 GGGGATGCCGCTGCTGGGTACTGCTGGGGGCGACGTGACAT----- 502  
QY 98 gLySaAnglyGly 102  
Db 503 ----AATGTGG 511

RESULT 33

LOCUS AR080463 3757 bp DNA linear PAT 31-AUG-2000  
DEFINITION Sequence 19 from patent US 5968782.  
ACCESSION AR080463  
VERSION AR080463.1 GI:10007198  
KEYWORDS  
ORGANISM Unknown.  
SOURCE Unknown.  
REFERENCE 1 (bases 1 to 3757)  
AUTHORS Stevens, R. L.  
TITLE Mast cell protease that cleaves fibrinogen  
JOURNAL Patent: US 5968782-A 19 19-OCT-1999;  
FEATURES  
source 1..3757  
location/Qualifiers  
1..3757  
/organism="Unknown"

BASE COUNT 802 a 1014 c 1003 g 938 t  
ORIGIN

Alignment Scores:

Pred. No.: 1.06 Length: 3757  
Score: 102.00 Matches: 21  
Percent Similarity: 62.50% Conservative: 9  
Best Local Similarity: 43.75% Mismatches: 18  
Query Match: 18.48% Indels: 0  
DB: 6 Gaps: 0

US-10-020-441-2 (1-106) x AR080463 (1-3757)

QY 46 GYpPheAspIleAlaIleValMetLeuAlaGlnMetValIleLeuGlnSerGlyIleArg 65  
Db 2419 GGGGCGAGAGCTTGGCTCTGGAGCTTGAAGTCCCTGTGATCTCCACCCCATATCCAC 2478  
QY 66 ValIleSerLeuProGlnProSerAspIleProProGlyIleArgValIleVal 85  
Db 2479 CCCATATCCCTGCGCCCTGCTGCGAGACCTTCCCTCGGAGACATCGTGGTGACA 2538  
QY 86 gLYTYrGlyArgAspAspAsp 93

Db 2539 GCGTGGGCGACATGATATATGAC 2562

MUSPROT6A

LOCUS MUSPROT6A 3757 bp mRNA linear ROD 27-APR-1993  
DEFINITION Mouse protease-6 gene, complete cds.  
ACCESSION M57625  
VERSION M57625.1 GI:200506  
KEYWORDS  
SOURCE Mouse (strain DBA/2) mast cell line K15V-MC, cDNA to mRNA.  
ORGANISM Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrate; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE

AUTHORS Reynolds, D.S., Gurley, D.S., Austen, K.F. and Serafin, W.E.  
TITLE Cloning of the cDNA and gene of mouse mast cell protease-6.  
Transcription by progenitor mast cells and mast cells of the  
connective tissue subclass  
JOURNAL J. Biol. Chem. 266 (6), 3847-3853 (1991)  
MEDLINE 91139682  
PubMed 1995638

FEATURES

source  
Location/Qualifiers  
1..3757  
/organism="Mus musculus"  
/strain="DBA/2"  
/db\_xref="taxon:10090"  
/cell\_line="K15V-MC"  
/cell\_type="mast"  
join(1572..1594,1803..1872,2012..2177,2238..2563,  
2697..2860,2934..3343)  
/gene="protease-6"  
2697..2860,2934..3343)  
/gene="protease-6"  
1572..1594  
/gene="protease-6"  
/number=1  
join(1803..1872,2012..2177,2238..2563,2697..2860,  
2934..3098)  
/gene="protease-6"  
/codon\_start=1  
/product="protease-6"  
/protein\_id="AAA39987.1"  
/db\_xref="GI:200507"  
/translation="MLKRRLLIMLALSLAIVSAPPANORVIGVGHSESKWP  
MQVSLRKLTNTYIRFCGSLHPQWLTFAICVGHITSPLFVQLPEOYLTYGDO  
LSLNIVVPHYITBEGADVALLEVPVAVSTHIFISLSPASSTPPTSCWYTG  
WGDIIDNDPPLPPYPLKQVKVIVENSICDRKHYTGIDPPIVHGMLCAGTFR  
DSCQDSGGLVCKVKGTWLDAGVSWGEGCAOPMKPGIYTRVLYLDMIRHYVBS  
"

exon <1803..1872  
/gene="protease-6"  
exon 2012..2177  
/gene="protease-6"  
exon 2238..2563  
/gene="protease-6"  
exon 2697..2860  
/gene="protease-6"  
exon 2934..3343  
/gene="protease-6"  
exon 2934..3343  
/gene="protease-6"  
BASE COUNT 802 a 1014 c 1003 g 938 t  
ORIGIN

Alignment Scores:

Pred. No.: 1.06 Length: 3757  
Score: 102.00 Matches: 21  
Percent Similarity: 62.50% Conservative: 9  
Best Local Similarity: 43.75% Mismatches: 18



Percent Similarity:	45.76%	Conservative:	21
Best Local Similarity:	27.97%	Mismatches:	36
Query Match:	18.39%	Indels:	29
DB:	9	Gaps:	4
US-10-020-441-2 (1-106) x AF318074 (1-2955)			
OY	6	LeuArgAnGlyAaPngInGlyIleHisIleGlnProSerGlyValIleAlaPro	25
Db	898	CTGGCTCCCGGGTGTCTCTGGGGGCTGCCAGGGCCCTGAATGGAACTCCCTGGCCA	957
OY	26	Gly-----TyrMetProSerCysMetSerAlaArgIleArgArg-----	38
Db	958	GGGACATCAAGAGATCTGGCGCCCTCAAGG-GTGCACATGGCGGAGACAGACCTTACTAC	1016
OY	39	-----ProIleAlaGlnThrIleu-----	44
Db	1017	CAGGACCGAGCTGCTGGCGGTGACGAGCATCTGTCGACCCACAGTTCTACATCATCCAG	1076
OY	45	SerGlyPheAspIleAlaIleValMetIleuAlaGlnMetValAsnLeuGlnSerGlyIle	64
Db	1077	ACCGGGGGGACATCTGCTCTGAGAGTGAGAGACCGCGTGAACATCTCCAGCCACATCT	1136
OY	65	ArgValIleSerLeuProGlnProSerAspIleProPropGlyThrGlyValPheIle	84
Db	1137	CACAGCGTACAGCTGCTCCCTGCTCGAGACACTTCCCCCGGGGAGATCCGCTGTGGCT	1196
OY	85	ValGlyTyrGlyArgAspAspAsnAspArgAspProSerArgLysAsnGlyGly	102
Db	1197	ACTGCTGGGGCGACGTGACAT-----AATGTGGG	1229
RESULT 38			
HS030A1		30228 bp	DNA
LOCUS			linear
DEFINITION			Human DNA sequence from clone LA16-303A1 on chromosome 16, complete
ACCESSION			AL031704
VERSION			AL031704.24
KEYWORDS			GT:11878023
SOURCE			HTG.
ORGANISM			human.
REFERENCE			Homo sapiens
AUTHORS			Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
TITLE			Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
JOURNAL			1 (bases 1 to 30228)
COMMENT			Lloyd, C.
			Direct Submission
			Submitted (16-DEC-2000) Sanger Centre, Hinxton, Cambridgeshire,
			CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk
			requests: clonerequests@sanger.ac.uk
			On Dec 17, 2000 this sequence version replaced gi:11230433.
			During sequence assembly data is compared from overlapping clones.
			Where differences are found these are annotated as variations
			together with a note of the overlapping clone name. Note that the
			variation annotation may not be found in the sequence submission
			corresponding to the overlapping clone, as we submit sequences with
			only a small overlap as described above.
			This sequence has been finished according to sequencing criteria
			as follows. An attempt is made to resolve all sequencing problems,
			such as compressions and repeats, but not necessarily within known
			annotated repeat sequence elements. Where the sequence is
			ambiguous, there is an annotation using the 'unsure' feature key.
			The following abbreviations are used to associate primary accession
			numbers given in the feature table with their source databases:
			Em; EMBL; Sw; SWISSPROT; Tr; TREMBL; Wp; WORMEP; Information
			on the WORMEP database can be found at
			http://www.sanger.ac.uk/Projects/C_elegans/wormep
			LA16-303A1 is
			part of a clone contig from the tip of the short arm of chromosome
			16 spanning 2Mb of p13.3 (Higgs D.R., Flint J., Daniels R., MRC
			Molecular Haematology Unit, Institute of Molecular Medicine, Oxford
			(unpublished), and is from the Los Alamos, flow sorted human
			Chromosome 16 libraries constructed by Norman Doggett
			(unpublished). VECTOR: scos-1
			IMPORTANT: This sequence is not the entire insert of clone



LA16-303A1 It may be shorter because we sequence overlapping sections only once, except for a 100 base overlap. The true left end of clone LA16-303A1 is at 1 in this sequence. The true left end of clone LA16-335B1 is at 30129 in this sequence. The true right end of clone LA16-337D8 is at 1812 in this sequence.

## FEATURES

source location/Qualifiers

1..30228

/organism="Homo sapiens"

/db\_xref="taxon:9606"

/chromosome="16"

/clone="LA16-303A1"

/clone\_11b="LA16"

3136..3197

/note="Weak data"

14153

/note="Random repeat. Forced join. Gap size estimated to be approximately 200bp by restriction digest data."

17718

/note="Random repeat. Forced join. Gap size estimated to be approximately 400bp by restriction digest data."

26789..26882

/note="Single clone region. In alu. Assembly confirmed by restriction digest data."

26904..26982

/note="Weak data in alu."

BASE COUNT 6183 a 8830 c 9037 g 6178 t

## ORIGIN

## Alignment Scores:

Pred. No.:

Score:

Percent Similarity:

Best Local Similarity:

Query Match:

DB:

US-10-020-441-2 (1-106) x HS303A1 (1-30228)

QY

DB

QY

DB

QY

DB

QY

DB

QY

DB

QY

DB

QY

DB

QY

DB

QY

DB

QY

DB

QY

DB

QY

DB

QY

DB

QY

DB

QY

DB

QY

DB

QY

REFERENCE 1 (bases 1 to 265786)  
AUTHORS Daniel, R.J., Peden, J.F., Lloyd, C., Horsley, S.W., Clark, K., Tufarelli, C., Kearney, L., Buckle, V.J., Doggett, N.A., Flint, J., and Higgs, D.R.  
TITLE Sequence, structure and pathology of the fully annotated terminal 2 Mb of the short arm of human chromosome 16  
JOURNAL Hum. Mol. Genet. 10 (4), 339-352 (2001)  
MEDLINE 21096910  
PUBMED 11157797  
REFERENCE 2 (bases 1 to 265786)  
AUTHORS Daniel, R.J., Peden, J.F., Lloyd, C., Horsley, S.W., Clark, K., Tufarelli, C., Kearney, L., Buckle, V.J., Doggett, N.A., Flint, J., and Higgs, D.R.  
TITLE Direct Submission  
JOURNAL Submitted (07-DEC-2000) MRC Molecular Haematology Unit, Weatherall Institute of Molecular Medicine, John Radcliffe Hospital, Oxford, Oxon OX3 9DS, UK

## FEATURES

source

source

source

source

source

source

source

source

source

source

source

source

source

source

source

source

source

source

source

source

source

source

source

source

source

source

source

source

source

source

source

source

source

source

source

source

source

source

source

source

source

source

source

source

source

source

source

source

source

1..265786  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="16"  
/map="16p13.3"  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="cosmid 305f3"  
14298..51556  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="cosmid 349e11"  
47200..79232  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="cosmid 381g6"  
68329..113551  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="cosmid 344f5"  
120839..154792  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="cosmid 320e12"  
140693..178475  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="cosmid 357d8"  
176599..212241  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="cosmid 303a1"  
206827..250642  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="cosmid 33e1"  
246926..265786  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="cosmid 358b7"  
1..1485  
/note="HS313F9"  
1..197  
/note="Alu0", RepeatMasker predicted  
23/10/2000  
/rpt\_family="SINE/Alu"  
complement(280..664)  
/note="MLTIC", RepeatMasker predicted  
23/10/2000  
/rpt\_family="LTR/MaLR"  
837..1937  
/note="Cpg island", not associated with any known gene"  
evidence=not\_experimental  
1386..14397

repeat_region	/note="HS305F3" 2205..2565 /note="MSTR1; RepeatMasker predicted 23/10/2000" /rpt_family="LTR/MaLR" 2567..2598 /note="(TTTG)n; RepeatMasker predicted 23/10/2000" /rpt_family="Simple_repeat" complement(2616..2506) /note="AluBg; RepeatMasker predicted 23/10/2000" /rpt_family="SINE/Alu" 2947..3240 /note="LTR32; RepeatMasker predicted 23/10/2000" /rpt_family="LTR/ERV_L" 3241..3337 /note="(TTG)n; RepeatMasker predicted 23/10/2000" /rpt_family="Simple_repeat" 3350..3515 /note="LTR32; RepeatMasker predicted 23/10/2000" /rpt_family="LTR/ERV_L" 4698..4748 /note="MIR; RepeatMasker predicted 23/10/2000" /rpt_family="SINE/MIR" complement(7543..7687) /note="MERSA; RepeatMasker predicted 23/10/2000" /rpt_family="DNA/MER1_type" 8207..8232 /note="(CCCCG)n; RepeatMasker predicted 23/10/2000" /rpt_family="Simple_repeat" 10855..10978 /note="AluDb; RepeatMasker predicted 23/10/2000" /rpt_family="SINE/Alu" 11335..11864 /note="Cr-rich; RepeatMasker predicted 23/10/2000" /rpt_family="Low_complexity" complement(12201..12427) /note="LINE3; RepeatMasker predicted 23/10/2000" /rpt_family="LINE/L1" complement(12994..13286) /note="AluDb; RepeatMasker predicted 23/10/2000" /rpt_family="SINE/Alu" 13787..15957 /note="CPG_island; SSTR5" /evidence=not_experimental 14298..47299 /note="HS349E11" 14397..15644 /gene="SSTR5" 14397..15644 /gene="SSTR5" /note="Similar to Unigene cluster Hs241375" 14432..15526 /gene="SSTR5" /note="PubMed:7908405" /codon_start=1 /product="somatostatin receptor type 5" /protein_id="AAK61266.1" /db_xref="GI:14336736" /db_xref="WMT:182455" /db_xref="SWISS-PROT:P35346" /translation="MEPLPASTPSMNAASPGASGGGDNRTLVGPAPSGARAVLV
---------------	------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------

```

VLVLVCAAGLGGNTLVITYVWLFPAKTKVTNTIYINLNAVADLVNMLGPELATONNA
SMPFEPVLICRLVMTLWDVNOQFVSCLTVMSVDRVLAVHPLSSARWRPRAKLAS
AAWVPLICLSWISLPLVADVOEGTCSNAPSEPVLMGAVFIITYALVGFAPDLVIC
LCVLLIVKRVAGVAVGCVRRSERSKRTVMVVLVFGCMVPEFTVNI VMLAVAL
POBPASGLYFVFVILSYVANSKANPVLYGLSNFROSFOKVLCLRKSGAKDADATE
LRFDRIROQDEATPPHRAANGLMOTSKL"
complement (18028, .18082)
/notes="L2; Repeatmasker predicted
23/10/2000"
/rpt family="LINE/L2"
19693, .19971
/notes="AluJo; Repeatmasker predicted
23/10/2000"
/rpt family="SINE/Alu"
complement (21110, .21550)
/notes="L1MB3; Repeatmasker predicted
23/10/2000"
/rpt family="LINE/L1"
complement (21580, .21658)
/notes="MLT1A1; RepeatMasker predicted
23/10/2000"
/rpt family="LTR/MaIR"
complement (21772, .22038)
/notes="AluSg; RepeatMasker predicted
23/10/2000"
/rpt family="SINE/Alu"
complement (22041, .22115)
/notes="MLT1A1; RepeatMasker predicted
23/10/2000"
/rpt family="LTR/MaIR"
complement (22117, .22164)
/notes="L1MB3; RepeatMasker predicted
23/10/2000"
/rpt family="LINE/L1"
complement (22218, .22495)
/notes="AluSx; RepeatMasker predicted
23/10/2000"
/rpt family="SINE/Alu"
complement (22578, .22860)
/notes="AluSp; RepeatMasker predicted
23/10/2000"
/rpt family="SINE/Alu"
23130, .23422
/notes="AluJo; RepeatMasker predicted
23/10/2000"

Alignment Scores:
Pred. No.: 87.1 Length: 265786
Score: 101.50 Matches: 33
Best Local Similarity: 45.76% Conservative: 21
Best Global Similarity: 27.97% Mismatches: 36
Query Match: 18.39% Indels: 29
DB: 9 Gaps: 4

US-10-020-441-2 (1-106) x AEO06466 (1-265786)

QY 6 LeuA-gAsnGlyAspGlnGlnGlyIleHisIsgInProSerGlyValIlyValAlaPro 25
Db 192416 CTGGCTCCCGGGGTGCTCTCGGGGGCTGCCACAGGGCCCTGAGTGGATCTCCGCGGCCCA 192475
QY 26 Gly-----TyMetProSerCySwetSerAlaArgGlnArgArg----- 38
Db 192476 GGGACATCAAGGATCTGGCCGCGCTCAGG-CTGCAACTCGCGGAGCAGCACTTACTAC 192534
QY 39 -----ProIleAlaGlnThrIleu----- 44
Db 192535 CAGGACCAAGCTGCTGGCTGACGAGATCATCTGTGCACCCACAAGTTTCAATCATCCAG 192594
QY 45 sArgGlyPheAspIleAlaIleValMetLeuAlaGlnMetValAsnLeuGlnSerGlyIle 64
Db 192595 ACCGGGGCGGAATCGCTCTGTGAGCTGGAGAGACCCGCTGAATCATCTCCAGCCACATC 192654
QY 65 ArgValIleSerIleuProGlnProSerAspIleProProProGlyThrGlyValPheIle 84

```

```

Db 192655 CACACGCTACGCTGCTCCCTGCTGAGACCTTCCCGGGGATGCCGTGCTGGCTC 192714
Oy 85 ValGlyTYrGlyArgAspAspAspArgSerArgLysAsnGlyGly 102
Db 192715 ACTGGCTGGGGGACGCTGACAT-----NATGGTGGG 192747

```

```

RESULT 40
AR080457 1219 bp DNA linear PAT 31-AUG-2000
LOCUS AR080457
DEFINITION Sequence 7 from patent US 5968782.
ACCESSION AR080457
VERSION AR080457.1 GI:10007192
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.

```

```

REFERENCE 1 (bases 1 to 1219)
AUTHORS Stevens,R.L.
TITLE Mast cell protease that cleaves fibrinogen
JOURNAL Patent: US 5968782-A 7 19-OCT-1999;
FEATURES
source 1..1219
location/Qualifiers

```

```

BASE COUNT 252 a 401 c 303 g 263 t
ORIGIN

```

```

Alignment Scores:
Pred. No.: 0.428 Length: 1219
Score: 101.00 Matches: 21
Percent Similarity: 60.78% Conservative: 10
Best Local Similarity: 41.18% Mismatches: 20
Query Match: 18.30% Indels: 0
DB: 6 Gaps: 0

```

US-10-020-441-2 (1-106) x AR080457 (1-1219)

```

Oy 43 ThrLeuSerGlyPheAspIleAlaIleValMetLeuAlaGlnMetValAsnLeuGlnSer 62
Db 528 ACCGAGATGGGGCGGACATCCCTACTTGAGCTCAGAACCCCTGTAAACATTTCACG 587
Oy 63 GlyIleArgValIleSerLeuProGlnProSerAspIleProProGlyThrGlyVal 82
Db 588 CATGTCCACCCCTGCTCCCTGCTCAGAGACCTTCCCTCAGGAACATTGTGC 647
Oy 83 PheIleValGlyTYrGlyArgAspAspAsp 93
Db 648 TGGGTGACAGCTGGGGAAACATCGACATGAT 680

```

Search completed: April 8, 2003, 17:11:20  
Job time : 2331.23 secs

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.4 p5 4578  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using SW model

Run on: April 9, 2003, 12:39:51 ; Search time 88.623 Seconds

(without alignments)  
159.378 Million cell updates/sec

Title: US-10-020-441-2

Perfect score: 552

Sequence: 1 VSFRLTNKNGDQGGIHQPSG.....YGRDDNDRDPKRNKGILKK 106

Scoring table: BLOSUM62

Gapop 10.0, Gapext 0.5

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :  
1: /SID52/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.\*  
2: /SID52/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.\*  
3: /SID52/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.\*  
4: /SID52/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.\*  
5: /SID52/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.\*  
6: /SID52/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.\*  
7: /SID52/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.\*  
8: /SID52/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.\*  
9: /SID52/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.\*  
10: /SID52/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.\*  
11: /SID52/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.\*  
12: /SID52/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.\*  
13: /SID52/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.\*  
14: /SID52/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.\*  
15: /SID52/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.\*  
16: /SID52/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.\*  
17: /SID52/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.\*  
18: /SID52/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.\*  
19: /SID52/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.\*  
20: /SID52/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.\*  
21: /SID52/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.\*  
22: /SID52/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.\*  
23: /SID52/gcgdata/geneseq/geneseq-emb1/AA2002.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	552	100.0	106	21	AAV67587
2	102	18.5	276	19	AAW64242
3	102	18.5	276	19	AAW63172
4	102	18.5	276	23	ABR05437
5	101.5	18.4	190	20	AAW68539
6	101.5	18.4	190	20	AAW68540
7	101	18.3	270	19	AAW64239
8	100	18.1	275	19	AAW64237
9	100	18.1	275	19	AAW63173
10	99	17.9	245	23	AAU12005

11	99	17.9	245	23	AAU12021	Human beta-II tryp
12	99	17.9	245	23	AAU12023	Human beta-II tryp
13	99	17.9	245	23	AAU12024	Human beta-II tryp
14	99	17.9	249	23	AAU12006	Human beta-I tryp
15	99	17.9	249	23	AAU12017	Human beta-II tryp
16	99	17.9	249	23	AAU12019	Human beta-II tryp
17	99	17.9	249	23	AAU12020	Human beta-II tryp
18	99	17.9	273	19	AAW64238	Human mast cell tr
19	99	17.9	273	19	AAW64235	Rat homologue of m
20	99	17.9	273	19	AAW63174	Human mast cell tr
21	99	17.9	691	23	AAU14348	Human proctase PRT
22	96	17.4	176	23	ABR57252	Mouse ischaemic co
23	95.5	17.3	242	7	AAW6060	Sequence of human
24	95.5	17.3	269	7	AAW61076	Human elastase I.
25	95.5	17.3	270	8	AAW60759	Human pancreas ela
26	95.5	17.3	270	8	AAW60759	Human pancreas ela
27	95.5	17.3	270	23	AAW67691	Human pancreatic c
28	93	16.8	273	21	AAW64243	Human lung mast ce
29	93	16.8	245	23	AAU12008	Recombinant human
30	93	16.8	245	23	AAU12013	Human beta-II tryp
31	93	16.8	245	23	AAU12015	Human beta-II tryp
32	93	16.8	245	23	AAU12016	Human beta-II tryp
33	93	16.8	249	21	AAV55011	Human beta-tryptas
34	93	16.8	249	23	AAU12007	Recombinant human
35	93	16.8	249	23	AAU12009	Human beta-II tryp
36	93	16.8	249	23	AAU12011	Human beta-II tryp
37	93	16.8	249	23	AAU12012	Human beta-II tryp
38	93	16.8	267	19	AAW64241	Human mast cell tr
39	93	16.8	267	19	AAW63176	Human mast cell tr
40	93	16.8	267	23	AAW64360	Protein TP52 diffie
41	93	16.8	274	19	AAW64240	Human mast cell tr
42	93	16.8	274	19	AAW63175	Human mast cell tr
43	92.5	16.5	242	7	AAW60061	Sequence of human
44	91	16.5	245	23	AAU12022	Human beta-II tryp
45	91	16.5	249	23	AAU12018	Human beta-II tryp

# ALIGNMENTS

RESULT 1	
AAV67587	AAV67587 standard; Protein; 106 AA.
AAV67587	
AC	AAV67587
XX	
DT	19-JUN-2000 (first entry)
XX	
DE	S. mansoni cercarial elastase protein fragment.
XX	
KW	Schistosoma parasite; fusion protein; cercarial elastase; vaccine;
XX	antibody response; schistosomicide
OS	Schistosoma mansoni.
XX	
PN	EP92582-A2.
XX	
PD	12-APR-2000.
XX	
PF	05-OCT-1999; 99EP-0307832.
XX	
PR	07-OCT-1998; 98GB-0021821.
XX	
PA	(UYWA-) UNIV WALES BANGOR.
XX	
PI	Doenhoff M, Sayers J;
XX	
DR	WPI: 2000-259136/23.
DR	N-PSDB; AA290608.
PT	New vaccine for treatment of Schistosoma infections contains a
PT	recombinant fusion protein comprising cercarial elastase sequence fused
PT	to bacterial, phage or viral protein -

XX Claim 3; Page 17; 26pp; English.  
PS  
XX The invention provides a vaccine comprising a recombinant fusion protein  
CC capable of eliciting immunity against Schistosoma parasites. The fusion  
CC protein comprises the 27 or 28 kDa cercarial elastase sequence of  
CC S. mansoni or an active fragment, homologue or variant, fused to a  
CC bacterial, phage or viral protein. The vaccine containing the fusion  
CC protein has been found to induce a significantly increased antibody  
CC response against schistosoma infections, compared to the use of S  
CC mansoni cercarial elastase in its native form. The present sequence  
CC represents the S. mansoni cercarial elastase protein fragment encoded by  
CC exon 2 of the cercarial elastase gene, used in the fusion protein of the  
CC invention.  
XX  
SQ Sequence 106 AA;  
Query Match 100.0%; Score 552; DB 21; Length 106;  
Best Local Similarity 100.0%; Pred. No. 1.2e-51;  
Matches 106; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
OY 1 VSFLRLRNGDQGHHSQGVKAPGYMPCMSARORPIAQTLSGFDIAIVMLAQMVL 60  
DB 1 VSFLRLRNGDQGHHSQGVKAPGYMPCMSARORPIAQTLSGFDIAIVMLAQMVL 60  
OY 61 QSGIRVLSLPQPSDIPPGTGVFIYVGYGRDNDPDSRKNGSILKK 106  
DB 61 QSGIRVLSLPQPSDIPPGTGVFIYVGYGRDNDPDSRKNGSILKK 106  
RESULT 2  
AAW64242  
ID AAW64242 standard; Protein; 276 AA.  
XX  
AC AAW64242;  
XX  
DT 24-NOV-1998 (first entry)  
XX  
DE Murine mast cell protease 6.  
XX  
KM MCP-6; mouse; mast cell protease 6; mMCP-7; tryptase-7;  
KW blood clot; anticoagulant; myocardial infarction; reocclusion;  
XX thromboembolism; cerebral embolism; thrombosis; therapy.  
XX  
OS Mus sp.  
XX  
FH Key Location/Qualifiers  
FT Peptide 1..21 /label= Sig\_peptide  
FT Peptide 22..31 /label= Pro\_peptide  
FT Protein 32..276 /label= Mat\_protein  
XX  
PN WO9824886-A1.  
XX  
PD 11-JUN-1998.  
XX  
PE 25-NOV-1997; 97WO-US21620.  
XX  
PR 04-DEC-1996; 96US-0032354.  
XX  
PA (BGHM ) BRIGHAM & WOMENS HOSPITAL.  
XX  
PI Stevens RL;  
XX  
DR WPI; 1998-333308/29.  
XX N-PSDB; AAV44332-33.  
PT New compositions containing tryptase-7, e.g. mouse mast cell  
PT protease-7 - are used to treat clot formation in e.g. myocardial  
PT infarction, reocclusion following angioplasty or pulmonary  
PT thromboembolism

XX Disclosure; Page 70-71; 92pp; English.  
PS  
XX This is the deduced amino acid sequence of mouse mast cell protease  
CC 6 (mMCP-6) (see also AAV44332-33). The invention provides:  
CC compositions comprising an isolated tryptase-7 (such as mMCP-7, see  
CC AAW64233); a method for treating a blood clot by administering a  
CC nucleic acid molecule that codes for a tryptase-7, or an expression  
CC product; a nucleic acid encoding a serine protease (SP); and a  
CC method of producing a mature SP. Tryptase-7 polypeptides can be  
CC used for treating disorders mediated by undesirable thrombus clot  
CC formation such as myocardial infarction and reocclusion following  
CC angioplasty of blood clots associated with pulmonary  
CC thromboembolism, deep vein thrombosis, cerebral embolism, renal  
CC vein and peripheral arterial thrombosis.  
XX  
SQ Sequence 276 AA;  
Query Match 18.5%; Score 102; DB 19; Length 276;  
Best Local Similarity 43.8%; Pred. No. 0.0078;  
Matches 21; Conservative 9; Mismatches 18; Indels 0; Gaps 0;  
OY 46 GFDIAIVMLAQMVLQSGIRVLSLPQPSDIPPGTGVFIYVGYGRDND 93  
DB 120 GADVALLELVVNVVSTHHPISLPASSTPPTGTCWVGSDIND 167  
RESULT 3  
AAW63172  
ID AAW63172 standard; Protein; 276 AA.  
XX  
AC AAW63172;  
XX  
DT 27-OCT-1998 (first entry)  
XX  
DE Mouse mast cell protease (mMCP-6) zymogen.  
XX  
KM Mast cell protease; MCP; mouse; inhibitor; peptide substrate; asthma;  
KW tryptase-6 protein; inflammatory disorder; allergic rhinitis; urticaria;  
KW antioedema; eczematous dermatitis; atopic dermatitis; anaphylaxis;  
KW hyperproliferative skin disease; peptic ulcer; hyperresponsiveness;  
XX inflammatory skin condition.  
XX  
OS Mus sp.  
XX  
PN WO9833812-A1.  
XX  
PD 06-AUG-1998.  
XX  
PE 30-JAN-1998; 98WO-US01865.  
XX  
PR 05-FEB-1997; 97US-0037090.  
XX  
PA (BGHM ) BRIGHAM & WOMENS HOSPITAL.  
XX  
PI Huang C, Stevens RL;  
XX  
DR WPI; 1998-437390/37.  
XX N-PSDB; AAV42708, AAV42709.  
PT Tryptase-6 complex inhibitory peptides - used to treat mast  
PT cell-mediated inflammatory disorders e.g. asthma  
XX  
PS Disclosure; Page 44; 69pp; English.  
XX  
CC This represents the mouse mast cell protease (mMCP-6) zymogen. The  
CC invention provides sequences shown in AAW63160 to AAW63169 that are  
CC inhibitors of mMCP-6. These peptides which are tryptase-6 complex  
CC inhibitors, can be used for treating a mast cell-mediated inflammatory  
CC disorder. The inhibitors can be used to treat inflammatory disorders  
CC including asthma, allergic rhinitis, urticaria and antioedema,  
CC eczematous dermatitis (atopic dermatitis), anaphylaxis,  
CC hyperproliferative skin disease, peptic ulcers, inflammatory bowel

CC disorder, hyperresponsiveness and inflammatory skin conditions.  
XX

Sequence 276 AA;

Query Match 18.5%; Score 102; DB 19; Length 276;  
Best Local Similarity 43.8%; Pred. No. 0.0078;

Matches 21; Conservative 9; Mismatches 18; Indels 0; Gaps 0;

QY 46 GFDIAIWMALQGMVNLQSGIRVLSLPQSDIPPQGTGVIVGYGRDND 93  
DB 120 GADVALLLEFVPAVNVSTHIIHISLPASSETPPGTSCHVWGMDIDND 167

RESULT 4  
ABB05437  
ID ABB05437 standard; Protein: 276 AA.

AC ABB05437;

DT 15-APR-2002 (first entry)

DE Mouse mast cell protease 6 precursor (MMP-6) SEQ ID NO:3.

KM Mouse; mast cell protease 6-like enzyme; enzyme; metastasis; asthma;  
KM tumour angiogenesis; atherosclerosis; neurodegenerative disease;  
KM pathogenic infection; antiasthmatic; antiarteriosclerotic; vasotropic;  
KM neuroprotective; osteoplastic; antitumour; cytostatic; antiinflammatory;  
KM inflammation; prion protein amyloid plaque; Creutzfeldt-Jakob disease;  
KM Genstmann-Strausler syndrome; Scrapie; reestenosis; osteoporosis.

OS Mus musculus.

PN W0200198340-A2.

PD 27-DEC-2001.

PF 21-JUN-2001; 2001WO-EP07002.

PR 23-JUN-2000; 2000US-213589P.

PA (FARB) BAYER AG.

PI Xiao Y;

DR WPI; 2002-147793/19.

PT Human mast cell protease 6-like enzyme polypeptide and polynucleotide  
for identifying modulators useful in treating diseases e.g. asthma,  
PT atherosclerosis, neurodegenerative disease and pathogenic infection -  
XX  
PS Disclosure; Fig 3; 88pp; English.

CC The present invention describes a human mast cell protease 6-like enzyme  
CC (I). (I) has antiasthmatic, antiarteriosclerotic, neuroprotective,  
CC vasoconstrictive, osteoplastic, antitumour, cytostatic and antiinflammatory  
CC activities. (I) is useful for the detection of a polynucleotide (II)  
CC which encodes (I), in a biological sample. (I) and (II) can be used for  
CC screening agents which decrease the activity of (I) or (II). (I), (II)  
CC and a reagent that modulates the activity of (I) or (II) can be used for  
CC treating a mast cell protease 6-like enzyme dysfunction related disease  
CC selected from metastasis of malignant cells, tumour angiogenesis,  
CC inflammation including asthma, atherosclerosis, neurodegenerative disease  
CC and pathogenic infection, where the neurodegenerative diseases include  
CC prion protein amyloid plaques of Genstmann-Strausler syndrome,  
CC Creutzfeldt-Jakob disease, and Scrapie, and other diseases such as  
CC reestenosis, and osteoporosis. The present sequence represents mouse mast  
CC cell protease 6 precursor (MMP-6), which is given in the exemplification  
CC of the present invention.

SO Sequence 276 AA;

Query Match 18.5%; Score 102; DB 23; Length 276;  
Best Local Similarity 43.8%; Pred. No. 0.0078;

Matches 21; Conservative 9; Mismatches 18; Indels 0; Gaps 0;

QY 46 GFDIAIWMALQGMVNLQSGIRVLSLPQSDIPPQGTGVIVGYGRDND 93  
DB 120 GADVALLLEFVPAVNVSTHIIHISLPASSETPPGTSCHVWGMDIDND 167

RESULT 5  
AAW68539  
ID AAW68539 standard; Protein: 190 AA.

AC AAW68539;

DT 12-FEB-1999 (first entry)

DE Amino acid sequence of the trypsin-L gene clone NVB352/1.

KM Tumour rejection antigen precursor; TRAP; TRH; leukaemia; screening;  
KM lymphoma; cancer; HLA; human lymphocyte antigen; vaccine;  
KM trypsin-L.

OS Homo sapiens.

PN W09849299-A1.

PD 05-NOV-1998.

PF 22-APR-1998; 98WO-US07784.

PR 25-APR-1997; 97US-0845998.

PA (LUDW-) LUDWIG INST CANCER RES.

PI Boon-Fallour T, Coule PG, De Smet C, Lucas S, Van Baren N;

DR WPI; 1999-009425/01.

DR N-PSDB; AAV33910.

PT New diagnosis of leukaemia - by detecting genes for tumour antigen  
rejection precursors or corresponding proteins

PS Claim 11; Pages 58-59; 88pp; English.

CC The present sequence represents the amino acid sequence of the  
CC trypsin-L gene clone NVB352/1. The protein is a tumour rejection  
CC antigen precursor (TRAP). The specification describes the treatment of  
CC disorders which characterised by expression of a leukaemia-associated  
CC nucleic acid such as TRH. The products are used for in vivo or in vitro  
CC screening for leukaemia, lymphoma or other cancers by usual  
CC hybridisation/amplification or immunoassay methods. TRAPs, when  
CC processed to antigens or complexed with HLA (human lymphocyte antigen)  
CC molecules, or nucleic acid encoding them, are useful in vaccines for  
CC treating leukaemia.

SO Sequence 190 AA;

Query Match 18.4%; Score 101.5; DB 20; Length 190;  
Best Local Similarity 39.7%; Pred. No. 0.0056;  
Matches 25; Conservative 10; Mismatches 19; Indels 9; Gaps 2;

QY 40 IAOITLSEFDIAIWMALQGMVNLQSGIRVLSLPQSDIPPQGTGVIVGYGRDND 99  
DB 115 IIOI--GADVALLLEFVPAVNVSTHIIHISLPASSETPPGTCMVWGMDVND-- 165

QY 100 NGG 102

DB 166 NGG 168

RESULT 6  
AAW68540  
ID AAW68540 standard; Protein: 190 AA.

AC AAW68540;  
 XX  
 DT 12-FEB-1999 (first entry)  
 XX  
 DE Amino acid sequence of the trypsin-L gene clone NVB352/3.  
 XX  
 KM Tumour rejection antigen precursor; TRAP; TRH; leukaemia; screening;  
 KM lymphoma; cancer; HLA; human lymphocyte antigen; vaccine;  
 KM trypsin-L.  
 XX  
 OS Homo sapiens.  
 XX  
 FN WO9849299-A1.  
 XX  
 PD 05-NOV-1998.  
 XX  
 PF 22-APR-1998; 98WO-US07784.  
 XX  
 PR 25-APR-1997; 97US-0845998.  
 XX  
 PA (LUDWIG INST CANCER RES.  
 XX  
 PI Boon-Falleur T, Coulie PG, De Smet C, Lucas S, Van Baren N;  
 XX  
 DR WPI; 1999-009425/01.  
 DR N-PSDB; AAV33911.  
 XX  
 PT New diagnosis of leukaemia - by detecting genes for tumour antigen  
 PT rejection precursors or corresponding proteins  
 XX  
 PS Claim 11; Page 62; 88pp; English.  
 XX  
 CC The present sequence represents the amino acid sequence of the  
 CC trypsin-L gene clone NVB352/3. The protein is a tumour rejection  
 CC antigen precursor (TRAP). The specification describes the treatment of  
 CC disorders which characterised by expression of a leukaemia-associated  
 CC nucleic acid such as TRH. The products are used for in vivo or in vitro  
 CC screening for leukaemia, lymphoma or other cancers by usual  
 CC hybridisation/amplification or immunoassay methods. TRAPs, when  
 CC processed to antigens or complexed with HLA (human lymphocyte antigen)  
 CC molecules, or nucleic acid encoding them, are useful in vaccines for  
 CC treating leukaemia.  
 CC  
 SQ Sequence 190 AA;  
 XX  
 Query Match 18.4%; Score 101.5; DB 20; Length 190;  
 Best Local Similarity 39.7%; Pred. No. 0.0056;  
 Matches 25; Conservative 10; Mismatches 19; Indels 9; Gaps 2;  
 QY 40 IAGTSGFDIAIVMLAQWNLQSGIRVISTLPQSDIPPGTGVFIVGYGRDNDPSPRK 99  
 DB 115 IIQT--GADIALLELPSPVNISSHIHTVLPASSTFPFGMPCWVTGMDVDN----- 165  
 QY 100 NGG 102  
 DB 166 NGG 168  
 XX  
 RESULT 7  
 ID AAW64239 standard; Protein; 270 AA.  
 AC AAW64239;  
 XX  
 DT 24-NOV-1998 (first entry)  
 XX  
 DE Gerbil homologue of mouse mMCp-7 zymogen.  
 XX  
 KM MCP-7 zymogen; gerbil; mast cell protease 7; trypsin-7;  
 KM blood clot; anticoagulant; myocardial infarction; reocclusion;  
 KM thromboembolism; cerebral embolism; thrombosis; therapy.  
 OS Meriones sp.

XX  
 FH Key Location/Qualifiers  
 FT Protein 25..270  
 FT /label= Mat\_protein  
 XX  
 PN WO9824886-A1.  
 XX  
 PD 11-JUN-1998.  
 XX  
 PF 25-NOV-1997; 97WO-US21620.  
 XX  
 PR 04-DEC-1996; 96US-0032354.  
 XX  
 PA (BGHM ) BRIGHAM & WOMENS HOSPITAL.  
 XX  
 PI Stevens RL;  
 PI  
 DR WPI; 1998-333308/29.  
 DR N-PSDB; AAV44326.  
 XX  
 PT New compositions containing trypsin-7, e.g. mouse mast cell  
 PT protease-7 - are used to treat clot formation in e.g. myocardial  
 PT infarction, reocclusion following angioplasty or pulmonary  
 PT thromboembolism  
 XX  
 PS Disclosure; Page 61; 92pp; English.  
 XX  
 CC This is the amino acid sequence of the gerbil homologue of mouse  
 CC mast cell protease 7 (mMCP-7, see also AAW64233), deduced from a  
 CC nucleic acid sequence (see AAW44326). Mature gerbil homologue of  
 CC mMCP-7 can be used to prevent or inhibit fibrin clot formation in  
 CC vitro or in vivo. The invention provides: compositions comprising  
 CC an isolated trypsin-7 such as mMCP-7, its homologues or humanised  
 CC mMCP-7; a method for treating a blood clot by administering a  
 CC nucleic acid molecule that codes for a trypsin-7, or an expression  
 CC product, to decrease fibrinogen activity; a nucleic acid encoding a  
 CC serine protease (SP); and a method of producing a mature SP by  
 CC expressing the inactive zymogen in a host cell, and cleaving the  
 CC enterokinase susceptibility domain. The trypsin-7 polypeptides  
 CC can be used to treat disorders mediated by undesirable thrombus  
 CC clot formation such as myocardial infarction and reocclusion  
 CC following angioplasty of blood clots associated with pulmonary  
 CC thromboembolism, deep vein thrombosis, cerebral embolism, renal  
 CC vein and peripheral arterial thrombosis. They are also useful for  
 CC all surgical procedures that require decreased blood clots.  
 CC  
 SQ Sequence 270 AA;  
 XX  
 Query Match 18.3%; Score 101; DB 19; Length 270;  
 Best Local Similarity 41.2%; Pred. No. 0.0097;  
 Matches 21; Conservative 10; Mismatches 20; Indels 0; Gaps 0;  
 QY 43 TLGFDIAIVMLAQWNLQSGIRVISTLPQSDIPPGTGVFIVGYGRDND 93  
 DB 111 TONGADIALLELPSPVNISSHIHTVLPASSTFPFGMPCWVTGMDVDND 161  
 XX  
 RESULT 8  
 ID AAW64237 standard; Protein; 275 AA.  
 AC AAW64237;  
 XX  
 DT 24-NOV-1998 (first entry)  
 XX  
 DE Human mast cell trypsin alpha.  
 XX  
 KM Mast cell trypsin alpha; human; MCP-7; mast cell protease 7;  
 KM blood clot; anticoagulant; myocardial infarction; reocclusion;  
 KM thromboembolism; cerebral embolism; thrombosis; therapy.  
 OS Homo sapiens.



```

FH Key Location/Qualifiers
FT Peptide 1..20
FT /label= Sig_peptide
FT Peptide 21..30
FT /label= Pro_peptide
FT Protein 31..275
FT /label= Mat_protein
PN WO9824886-A1.
XX 11-JUN-1998.
XX 25-NOV-1997; 97WO-US21620.
XX 04-DEC-1996; 96US-0032354.
XX (BGHM ) BRIGHAM & WOMENS HOSPITAL.
XX Stevens RL;
PI WPI; 1998-333308/29.
DR N-PSDB; AAV44328.
XX
XX New compositions containing tryptase-7, e.g. mouse mast cell
PT protease-7 - are used to treat clot formation in e.g. myocardial
PT infarction, reocclusion following angioplasty or pulmonary
PT thromboembolism
XX
XX Disclosure; Page 63-64; 92pp; English.
PS
XX This is the deduced amino acid sequence of human mast cell tryptase
CC alpha (see also AAV44328). The invention provides: compositions
CC comprising an isolated tryptase-7 that may include chimeric proteins
CC that contain (a) a human tryptase-7 for all but the active site region
CC and (b) the substrate-binding pocket of mouse tryptase-7 or its
CC homologues (see AAW64233-39); a method for treating a blood clot by
CC administering a nucleic acid molecule that codes for a tryptase-7,
CC or an expression product, to decrease fibrinogen activity; a nucleic
CC acid encoding a serine protease (SP); and a method of producing a
CC mature SP by expressing the inactive zymogen in a host cell, and
CC cleaving the enterokinase susceptibility domain. The tryptase-7
CC polypeptides can be used to treat disorders mediated by undesirable
CC thrombus clot formation such as myocardial infarction and
CC reocclusion following angioplasty of blood clots associated with
CC pulmonary thromboembolism, deep vein thrombosis, cerebral embolism,
CC renal vein and peripheral arterial thrombosis. They are also
CC useful for all surgical procedures that require decreased blood
CC clots.
XX
SQ Sequence 275 AA;
Query Match 18.1%; Score 100; DB 19; Length 275;
Best Local Similarity 40.7%; Pred. No. 0.013;
Matches 22; Conservative 10; Mismatches 20; Indels 2; Gaps 1;
QY 40 IAQTLSGFDIAVMAQWVNLQSGIRVLSLPQSPDIPPGTGVFVGYGRDND 93
| | | | | | | | | | | | | | | | | | | | | | | | | |
| | | | | | | | | | | | | | | | | | | | | | | | | |
Dd 115 IIQI--GADIALLEBPVNISRVHTVMLPPASFTPPGMCVWTGWDVND 166

```

```

KW hyperproliferative skin disease; peptic ulcer; hyperresponsiveness;
KW inflammatory skin condition; human; mast cell tryptase alpha.
XX
XX Homo sapiens.
OS
XX WO9833812-A1.
XX
XX 06-AUG-1998.
XX
XX 30-JAN-1998; 98WO-US01865.
XX
XX 05-FEB-1997; 97US-0037090.
XX
XX (BGHM ) BRIGHAM & WOMENS HOSPITAL.
XX
XX Huang C, Stevens RL;
PI WPI; 1998-437390/37.
DR N-PSDB; AAV42710.
XX
XX Tryptase-6 complex inhibitory peptides - used to treat mast
PT cell-mediated inflammatory disorders e.g. asthma
XX
XX Disclosure; Pages 45-46; 69pp; English.
PS
XX This represents the human mast cell tryptase alpha protein which is a
CC homologue of the mouse mast cell protease (mMCP-6) zymogen. The
CC invention provides sequences shown in AAW63160 to AAW63169 that are
CC inhibitors of mMCP-6. These peptides which are tryptase-6 complex
CC inhibitors, can be used for treating a mast cell-mediated inflammatory
CC disorder. The inhibitors can be used to treat inflammatory disorders
CC including asthma, allergic rhinitis, urticaria and anioedema,
CC eczematous dermatitis (atopic dermatitis), anaphylaxis,
CC hyperproliferative skin disease, peptic ulcers, inflammatory bowel
CC disorder, hyperresponsiveness and inflammatory skin conditions.
XX
SQ Sequence 275 AA;
Query Match 18.1%; Score 100; DB 19; Length 275;
Best Local Similarity 40.7%; Pred. No. 0.013;
Matches 22; Conservative 10; Mismatches 20; Indels 2; Gaps 1;
QY 40 IAQTLSGFDIAVMAQWVNLQSGIRVLSLPQSPDIPPGTGVFVGYGRDND 93
| | | | | | | | | | | | | | | | | | | | | | | | | |
| | | | | | | | | | | | | | | | | | | | | | | | | |
Dd 115 IIQI--GADIALLEBPVNISRVHTVMLPPASFTPPGMCVWTGWDVND 166

```

XX		WT; 2002-11A578/15.
DR	N-PSDB; AAS20760.	
XX		
PT	DNA construct for producing enzymatically-inactive proteolytic	
PT	tryptase, comprises DNA sequence encoding proteolytic tryptase having	
PT	an active site mutation -	
XX		
PS	Disclosure; Page 71-72; 126pp; English.	
XX		
CC	The present invention relates to recombinant human proteolytic	
CC	tryptases, active site mutants of these trypsinases and the methods for	
CC	producing these. The method involves the production of a DNA expression	
CC	construct comprising a promoter operably linked to a secretion signal	
CC	sequence which is operably linked to a DNA sequence encoding a	
CC	proteolytic tryptase with an active site mutation (the construct drives	
CC	expression of a mature proteolytic tryptase that lacks enzymatic activity	
CC	due to the active site mutation, in hosts transformed to contain the	
CC	construct). The method is useful for producing enzymatically-active	
CC	beta-II tryptase. The active site mutants of proteolytic tryptase provide	
CC	a tool to investigate the structural and functional properties of the	
CC	protease and its enzymatic activity, and for modelling studies. The	
CC	enzymatically-active, recombinant proteolytic tryptase produced are	
CC	useful as an antigen to generate anti-human tryptase antibodies	
CC	and in drug screening for compounds which act as tryptase inhibitors,	
CC	antagonists, agonists, etc. The present sequence represents human	
CC	mature beta-I tryptase.	
SQ	Sequence 245 AA;	
	Query Match 17.9%; Score 99; DB 23; Length 245;	
	Best Local Similarity 39.6%; Pred. No. 0.014;	
	Matches 19; Conservative 11; Mismatches 18; Indels 0; Gaps 0;	
Oy	46 GFDDIWMIAQWVNLQSGIRIVSLPPGPIPGGVFIYGGDDND 93	
	:   :   :   :   :   :   :   :   :   :   :	
Dd	89 GADIDALELEEFVNVSHTVTITLPPASETFPPGHPCWVGMDVDND 136	
RESULT 11		
ID	AAU12021 standard; Protein; 245 AA.	
XX	AAU12021;	
AC		
XX	09-APR-2002 (first entry)	
DT		
XX	Human beta-II tryptase active site mutant H44A #4.	
DE		
XX	Human; proteolytic tryptase; protease; recombinant beta-II tryptase;	
KW	enzyme; mutant; mutein.	
RW		
XX	Homo sapiens.	
OS	Synthetic.	
OS		
XX	WO200198470-A2.	
PN		
XX	27-DEC-2001.	
PD		
XX	20-JUN-2001; 2001WO-US19681.	
PF		
PR	21-JUN-2000; 2000US-0598982.	
XX	(PROM-) PROMEGA CORP.	
PA		
XX	Maffit M, Niles AL, Haak-Frendescho M;	
F1		
XX	WT; 2002-11A578/15.	
DR	N-PSDB; AAS20787.	
XX		
PT	DNA construct for producing enzymatically-inactive proteolytic	
PT	tryptase, comprises DNA sequence encoding proteolytic tryptase having	
PT	an active site mutation -	

XX Claim 40; Page 117-118; 126pp; English.

PX

CC The present invention relates to recombinant human proteolytic  
CC trypsinases, active site mutants of these trypsinases and the methods for  
CC producing these. The method involves the production of a DNA expression  
CC construct comprising a promoter operably linked to a secretion signal  
CC sequence which is operably linked to a DNA sequence encoding a  
CC proteolytic trypsinase with an active site mutation (the construct drives  
CC expression of a mature proteolytic trypsinase that lacks enzymatic activity  
CC due to the active site mutation, in hosts transformed to contain the  
CC construct). The method is useful for producing enzymatically-active  
CC beta-II trypsinase. The active site mutants of proteolytic trypsinase provide  
CC a tool to investigate the structural and functional properties of the  
CC protease and its enzymatic activity, and for modelling studies. The  
CC enzymatically-active, recombinant proteolytic trypsinase produced are  
CC useful as an antigen to generate anti-human trypsinase antibodies  
CC and in drug screening for compounds which act as trypsinase inhibitors,  
CC antagonists, agonists, etc. AAU1009-AAU12024 represent recombinant  
CC human beta-II trypsinase active site mutants.  
CX XX

SQ Sequence 245 AA;

Query Match 17.9%; Score 99; DB 23; Length 245;  
Best Local Similarity 39.6%; Pred. No. 0.014;  
Matches 19; Conservative 11; Mismatches 18; Indels 0; Gaps 0.

OY 46 GFDAIIVMLACWNVNLSGIRVSLPSPSDIPPPGTGVFIVEGYGRDND 93  
|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|  
Db 69 GADIDILBELPEPVNSHVHTVLTPPASETPPPGKPCWVTGMGDVDND 136

RESULT 12  
AAU12023  
ID AAU12023 standard; Protein; 245 AA.

AC AAU12023;  
XX  
DT 09-APR-2002 (first entry)  
XX  
DE Human beta-II trypsinase active site mutant SI94A #7.  
XX  
KM Human; proteolytic trypsinase; proenzyme; recombinant beta-II trypsinase;  
XX enzyme; mutant; mutein.  
XX  
OS Homo sapiens.  
OS Synthetic.  
PN WO200198470-A2.  
XX  
PD 27-DEC-2001.  
XX  
PF 20-JUN-2001; 2001WO-US19681.  
XX  
PR 21-JUN-2000; 2000US-0598982.  
XX  
PA (PROM-) PROMEGA CORP.  
XX  
PI Maffei M, Niles AL, Haak-Frendscho M;  
XX  
DR WPI; 2002-114578/15.  
XX  
DR N-PsDB; AAS20789.  
XX  
PT DNA construct for producing enzymatically-inactive proteolytic  
PT trypsinase, comprises DNA sequence encoding proteolytic trypsinase having  
PT an active site mutation -  
XX  
PS  
PS Claim 40; Page 122-123; 126pp; English.

CC The present invention relates to recombinant human proteolytic  
CC trypsinases, active site mutants of these trypsinases and the methods for  
CC producing these. The method involves the production of a DNA expression  
CC construct comprising a promoter operably linked to a secretion signal



Query Match 17.9%; Score 99; DB 23; Length 249;  
 Best Local Similarity 39.6%; Pred. No. 0.014;  
 Matches 19; Conservative 11; Mismatches 18; Indels 0; Gaps 0.

RESULT 15  
AAU12017  
ID AAU12017 standard; Protein; 249 AA

DT	09-APR-2002 (first entry)
XX	
XX	
DE	Human beta-II tryptase active site mutant H44A #3.
XX	
KM	Human; proteolytic tryptase; protease; recombinant beta-II tryptase;
KM	enzyme; mutant; mutein.

OS	Homo sapiens.
OS	Synthetic.
XX	
PN	WO200198470-A2.

PD 27-DEC-2001.

PF 20-JUN-2001; 2001WO-US19681.

PR 21-JUN-2000; 2000US-0598982.

PA (PROM-) PROMEGA CORP.

PI Maffit M, Niles AL, Haak-Frendscho M;

DR WPI; 2002-114578/15.

DR N-PSDB; AAS20783.

PT DNA construct for producing enzymatically-inactive proteolytic  
PT trypsin, comprises DNA sequence encoding proteolytic trypsin having  
PT an active site mutation -

PS Claim 8; Page 105-106; 126pp; English.

The present invention relates to recombinant human proteolytic trypsinases, active site mutants of these trypsinases and the methods for producing these. The method involves the production of a DNA expression construct comprising a promoter operably linked to a secretion signal sequence which is operably linked to a DNA sequence encoding a proteolytic trypsinase with an active site mutation. The construct drives expression of a mature proteolytic trypsinase that lacks enzymatic activity due to the active site mutation, in hosts transformed to contain the construct. The method is useful for producing enzymatically active beta-II trypsinase. The active site mutants of proteolytic trypsinase provide a tool to investigate the structural and functional properties of the proenzyme and its enzymatic activity, and for modelling studies. The enzymatically-active, recombinant proteolytic trypsinase produced are useful as an antigen to generate anti-human trypsinase antibodies and in drug screening for compounds which act as trypsinase inhibitors, antagonists, agonists, etc. AAU2009-AAU2024 represent recombinant human beta-II trypsinase active site mutants.

**SQ Sequence 249 AA;**

Query Match	17.9%	Score 99;	DB 23;	Length 249;
Best Local Similarity	39.6%	Pred. No. 0.014;		
Matches 19; Conservative	11;	Mismatches 18;	Indels 0;	Gaps 0

QY 46 GFDAIWMLAQMVNLQSGIRVLSLPQSPDIPPPGTGVFIWGYGRDND 93

Db 93 GADIALELEP VNVSSHVHTVTLPPASETFPPGMPCWVTGWDVDND 140

RESULT 16  
AAU12019  
ID AAU12019 standard; Protein; 249 AA

DT 09-APR-2002 (first entry)

DE Human beta-II tryptase active site mutant S194A #5.

**KW** Human; proteolytic triptase; protease; recombinant beta-II triptase;  
**KW** enzyme; mutant; mutein.

OS	Homo sapiens.
OS	Synthetic.

PN WO200198470-A2.

PD 27-DEC-2001.

20-JUN-2001; 2001WO-US196681.

PR 21-JUN-2000; 2000US-0598982.

PA (PROM-) PROMEGA CORP.

PI Maffit M, Niles AL, Haak-Frendscho M

DR WPI; 2002-114578/15.

DR N-PSDB; AAS20785.

PT DNA construct for producing enzymatically-inactive proteolytic trypsinase, comprises DNA sequence encoding proteolytic trypsinase having

XX

PS Claim 8; Page 111-112; 126pp; English.

The present invention relates to recombinant human proteolytic triptases, active site mutants of these triptases and the methods for producing these. The method involves the production of a DNA expression construct comprising a promoter operably linked to a secretion signal sequence which is operably linked to a DNA sequence encoding a proteolytic triptase with an active site mutation (the construct drives expression of a mature proteolytic triptase that lacks enzymatic activity due to the active site mutation, in hosts transformed to contain the construct). The method is useful for producing enzymatically-active beta-II triptase. The active site mutants of proteolytic triptase provide a tool to investigate the structural and functional properties of the proenzyme and its enzymatic activity, and for modelling studies. The enzymatically-active, recombinant proteolytic triptase antibodies are useful as an antigen to generate anti-human triptase antbodies and in drug screening for compounds which act as triptase inhibitors, antagonists, agonists, etc. AAU1209-AAU1204 represent recombinant human beta-II triptase active site mutants.

Sequence 249 AA;

Query Match	17.9%	Score 99;	DB 23;	Length 249;
Best Local Similarity	39.6%	Pred. No. 0.014;		
Matches 19; Conservative	11;	Mismatches 18;	Indels 0;	Gaps 0;

QY 46 GFDIAI VMLAQM VNL OSGIRVISL PQPSDI PPPTGV FIVGYGRDDND 93

Db 93 GADIALEEEPVNVSSHVHTVTLPPASETFPPGMPCWVTGWDVDND 140

## RESULT 17

ID AAU12020 standard; Protein; 249 AA

XXX

AAU12020;  
09-APR-2002 (first entry)  
Human beta-II tryptase active site mutant S194A #5.  
Human; proteolytic tryptase; protease; recombinant beta-II tryptase;  
enzyme; mutant; mutein.  
Homo sapiens.  
Synthetic.  
W0200198470-A2.  
27-DEC-2001.  
20-JUN-2001; 2001WO-US19681.  
21-JUN-2000; 2000US-0598982.  
(PROM-) PROMEGA CORP.  
Maffic M, Niles AL, Haak-Frendescho M;  
WPI; 2002-114578/15.  
N-PSDB; AAS20786.  
DNA construct for producing enzymatically-inactive proteolytic  
tryptase, comprises DNA sequence encoding proteolytic tryptase having  
an active site mutation -  
Claim 8; Page 114-115; 126BP; English.

The present invention relates to recombinant human proteolytic  
tryptases, active site mutants of these tryptases and the methods for  
producing these. The method involves the production of a DNA expression  
construct comprising a promoter operably linked to a secretion signal  
sequence which is operably linked to a DNA sequence encoding a  
proteolytic tryptase with an active site mutation (the construct drives  
expression of a mature proteolytic tryptase that lacks enzymatic activity  
due to the active site mutation, in hosts transformed to contain the  
construct). The method is useful for producing enzymatically-active  
beta-II tryptase. The active site mutants of proteolytic tryptase provide  
a tool to investigate the structural and functional properties of the  
protease and its enzymatic activity, and for modelling studies. The  
enzymatically-active, recombinant proteolytic tryptase produced are  
useful as an antigen to generate anti-human tryptase antibodies  
and in drug screening for compounds which act as tryptase inhibitors,  
antagonists, agonists, etc. AAU12009-AAU12024 represent recombinant  
human beta-II tryptase active site mutants.

Sequence 249 AA;

Query Match 17.9%; Score 99; DB 23; Length 249;  
Best Local Similarity 39.6%; Pred. No. 0.014;  
Matches 19; Conservative 11; Mismatches 18; Indels 0; Gaps 0;

46 GFDIAIWLAMQVNLQSGIRVISPQSDIPPGTGVIVGYGRDND 93  
93 GADIALLELEBPVNVSSHVTITLPPASERFPFGMPCWVGWDVND 140

RESULT 18  
AAW64238  
ID AAW64238 standard; Protein; 273 AA.

AAW64238;  
24-NOV-1998 (first entry)  
Human mast cell tryptase I.  
Mast cell tryptase I; human; MCP-7; mast cell protease 7;

blood clot; anticoagulant; myocardial infarction; reocclusion;  
thromboembolism; cerebral embolism; thrombosis; therapy.  
Homo sapiens.  
Key Location/Qualifiers  
Peptide 1..18  
Peptide /label= Sig\_peptide  
Peptide 19..28  
Protein /label= Pro\_peptide  
Protein 29..273  
/label= Mat\_protein  
W09824886-A1.  
11-JUN-1998.  
25-NOV-1997; 97WO-US21620.  
04-DEC-1996; 96US-0032354.  
(BGHM ) BRIGHAM & WOMENS HOSPITAL.  
Stevens RL;  
WPI; 1998-333308/29.  
N-PSDB; AAV44329.  
New compositions containing tryptase-7, e.g. mouse mast cell  
protease-7 - are used to treat clot formation in e.g. myocardial  
infarction, reocclusion following angioplasty or pulmonary  
thromboembolism  
Disclosure; Page 65; 92pp; English.

This is the deduced amino acid sequence of human mast cell tryptase  
I (see also AAV44329). The invention provides: compositions  
comprising an isolated tryptase-7 that may include chimeric proteins  
that contain (a) a human tryptase for all but the active site region  
and (b) the substrate-binding pocket of mouse tryptase-7 or its  
homologues (see AAW64233-39); a method for treating a blood clot by  
administering a nucleic acid molecule that codes for a tryptase-7,  
or an expression product, to decrease fibrinogen activity; a nucleic  
acid encoding a serine protease (SP); and a method of producing a  
mature SP by expressing the inactive zymogen in a host cell, and  
cleaving the enterokinase susceptibility domain. The tryptase-7  
polypeptides can be used to treat disorders mediated by undesirable  
thrombus clot formation such as myocardial infarction and  
reocclusion following angioplasty of blood clots associated with  
pulmonary thromboembolism, deep vein thrombosis, cerebral embolism,  
renal vein and peripheral arterial thrombosis. They are also  
useful for all surgical procedures that require decreased blood  
clots.

Sequence 273 AA;

Query Match 17.9%; Score 99; DB 19; Length 273;  
Best Local Similarity 39.6%; Pred. No. 0.016;  
Matches 19; Conservative 11; Mismatches 18; Indels 0; Gaps 0;

46 GFDIAIWLAMQVNLQSGIRVISPQSDIPPGTGVIVGYGRDND 93  
117 GADIALLELEBPVNVSSHVTITLPPASERFPFGMPCWVGWDVND 164

RESULT 19  
AAW64235  
ID AAW64235 standard; Protein; 273 AA.

AAW64235;  
24-NOV-1998 (first entry)

DE Rat homologue of mouse mMCP-7 zymogen.  
 XX  
 KM MCP-7 zymogen; rat; mast cell protease 7; tryptase-7;  
 KM blood clot; anticoagulant; myocardial infarction; reocclusion;  
 KM thromboembolism; cerebral embolism; thrombosis; therapy.  
 XX  
 OS Rattus sp.  
 XX  
 FH Key Location/Qualifiers  
 FT Protein 28..273  
 FT /label= Mat\_protein  
 XX  
 PN WO9824886-A1.  
 XX  
 PD 11-JUN-1998.  
 XX  
 PF 25-NOV-1997; 97WO-US21620.  
 XX  
 XX 04-DEC-1996; 96US-0032354.  
 XX  
 XX (BGHM ) BRIGHAM & WOMENS HOSPITAL.  
 XX  
 PI Stevens RL;  
 XX  
 DR WPI; 1998-333308/29.  
 XX  
 PT New compositions containing tryptase-7, e.g. mouse mast cell  
 PT protease-7 - are used to treat clot formation in e.g. myocardial  
 PT infarction, reocclusion following angioplasty or pulmonary  
 PT thromboembolism  
 XX  
 PS Disclosure; Page 59-60; 92pp; English.  
 XX  
 CC This is the amino acid sequence of the rat homologue of mouse  
 CC mast cell protease 7 (mMCP-7, see also AAM64233). Mature rat mMCP-7  
 CC homologue can be used to prevent or inhibit fibrin clot formation in  
 CC vitro or in vivo. The invention provides: compositions comprising  
 CC an isolated tryptase-7 such as mMCP-7, its homologues or humanised  
 CC mMCP-7; a method for treating a blood clot by administering a  
 CC nucleic acid molecule that codes for a tryptase-7, or an expression  
 CC product, to decrease fibrinogen activity; a nucleic acid encoding a  
 CC serine protease (SP); and a method of producing a mature SP by  
 CC expressing the inactive zymogen in a host cell, and cleaving the  
 CC enterokinase susceptibility domain. The tryptase-7 polypeptides  
 CC can be used to treat disorders mediated by undesirable thrombus  
 CC clot formation such as myocardial infarction and reocclusion  
 CC following angioplasty of blood clots associated with pulmonary  
 CC thromboembolism, deep vein thrombosis, cerebral embolism, renal  
 CC vein and peripheral arterial thrombosis. They are also useful for  
 CC all surgical procedures that require decreased blood clots.  
 XX  
 SO Sequence 273 AA;  
 XX  
 Query Match 17.9%; Score 99; DB 19; Length 273;  
 Best Local Similarity 40.7%; Pred. No. 0.016;  
 Matches 22; Conservative 10; Mismatches 20; Indels 2; Gaps 1;  
 XX  
 OY 40 IAQTLSGDIAVLAQMWNLOSIRVSLPPSPDIPPGNGVIFVYGRDND 93  
 DB 113 IAQ--DGADIALKLTNPVNITNSVHTVSLPPASSTFPSPGTLCTWVTGMCINND 164  
 XX  
 RESULT 20  
 AAM63174  
 ID AAM63174 standard; Protein; 273 AA.  
 XX  
 AC AAM63174;  
 XX  
 DT 27-OCT-1998 (first entry)  
 XX  
 DE Human mast cell tryptase I polypeptide.  
 XX  
 KM Mast cell protease; MCP; mouse; inhibitor; peptide substrate; asthma;

KM tryptase-6 protein; inflammatory disorder; allergic rhinitis; urticaria;  
 KM antioedema; eczematous dermatitis; atopic dermatitis; anaphylaxis;  
 KM hyperproliferative skin disease; peptic ulcer; hyperresponsiveness;  
 KM inflammatory skin condition; human; mast cell tryptase I.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO9833812-A1.  
 XX  
 PD 06-AUG-1998.  
 XX  
 PF 30-JAN-1998; 98WO-US01865.  
 XX  
 XX 05-FEB-1997; 97US-0037090.  
 XX  
 XX (BGHM ) BRIGHAM & WOMENS HOSPITAL.  
 XX  
 PI Huang C, Stevens RL;  
 XX  
 DR WPI; 1998-437390/37.  
 DR N-PSDB; AAV42711.  
 XX  
 PT Tryptase-6 complex inhibitory peptides - used to treat mast  
 PT cell-mediated inflammatory disorders e.g. asthma  
 XX  
 PS Disclosure; Pages 46-47; 69pp; English.  
 XX  
 CC This represents the human mast cell tryptase I protein which is a  
 CC homologue of the mouse mast cell protease (mMCP-6) zymogen. The  
 CC invention provides sequences shown in AAM63160 to AAM63169 that are  
 CC inhibitors of mMCP-6. These peptides which are tryptase-6 complex  
 CC inhibitors, can be used for treating a mast cell-mediated inflammatory  
 CC disorder. The inhibitors can be used to treat inflammatory disorders  
 CC including asthma, allergic rhinitis, urticaria and antioedema,  
 CC eczematous dermatitis (atopic dermatitis), anaphylaxis,  
 CC hyperproliferative skin disease, peptic ulcers, inflammatory bowel  
 CC disorder, hyperresponsiveness and inflammatory skin conditions.  
 XX  
 SO Sequence 273 AA;  
 XX  
 Query Match 17.9%; Score 99; DB 19; Length 273;  
 Best Local Similarity 39.6%; Pred. No. 0.016;  
 Matches 19; Conservative 11; Mismatches 18; Indels 0; Gaps 0;  
 XX  
 OY 46 GFDAIVLAQMWNLOSIRVSLPPSPDIPPGNGVIFVYGRDND 93  
 DB 117 GADIALLESPVNVSSVHTVTLPPASSTFPSPGTLCTWVTGMDVND 164  
 XX  
 RESULT 21  
 AAE14348  
 ID AAE14348 standard; Protein; 691 AA.  
 XX  
 AC AAE14348;  
 XX  
 DT 07-MAR-2002 (first entry)  
 XX  
 DE Human protease PRTS-13 protein.  
 XX  
 KM Human; protease; PRTS-13; tranquilliser; gene therapy; vaccine; allergy;  
 KM infection; dermatitis; arteriosclerosis; rheumatoid arthritis; hepatitis;  
 KM atherosclerosis; psoriasis; Alzheimer's disease; mental disorder; cancer;  
 KM gastrointestinal disorder; Cushing's syndrome; seizure; glaucoma; stroke;  
 KM epithelial disorder; urticaria; anorexia; trauma; asthma; eczema; nausea;  
 KM hypertension; neurological disorder; Parkinson's disease; drug screening;  
 KM cardiac; cell proliferative disorder; multiple sclerosis; osteoporosis;  
 KM diabetes mellitus; glomerulonephritis; cardiovascular disorder; anaemia;  
 KM autoimmune disorder; inflammatory disorder; myocardial infarction; AIDS;  
 KM developmental disorder; reproductive disorder; infertility; diarrhoea;  
 KM dementia; acidosis; cataract; gynaecomastia; epilepsy; jaundice.  
 XX  
 OS Homo sapiens.  
 XX

Key Location/Qualifiers  
 Peptide 1..28  
 label= Signal\_peptide  
 Protein 29..691  
 /note= "Mature human protease PRS-13 protein"  
 Active-site 65..108  
 Domain 67..80  
 /note= "Type I fibronectin domain"  
 Active-site 213..258  
 Domain 225..238  
 /note= "Type I fibronectin domain"  
 Domain 371..389  
 /note= "Transmembrane domain"  
 Active-site 425..473  
 Domain 433..450  
 /note= "Kring domain protein"  
 Domain 497..517  
 /note= "Transmembrane domain"  
 Domain 514..535  
 /note= "Kring domain protein"  
 Active-site 574..618  
 Domain 594..635  
 /note= "Kring domain protein"  
 Domain 604..638  
 /note= "Type I fibronectin domain"  
 WO200183775-A2.  
 08-NOV-2001.  
 04-MAY-2001; 2001WO-US14651.  
 04-MAY-2000; 2000US-202082P.  
 11-MAY-2000; 2000US-203566P.  
 17-MAY-2000; 2000US-205803P.  
 25-MAY-2000; 2000US-207477P.  
 01-JUN-2000; 2000US-209402P.  
 (INCY-) INCYTE GENOMICS INC.  
 Deleagane AM, Lal P, Hafalia A, Patterson C, Walla NK, Kearney L,  
 Tribouley CM, Khan FA, Yao WG, Baughn KR, Azimzai Y, Elliott VS,  
 Nguyen DB, Gandhi AR, Yang J, Hernandez R, Policky JL, Lu DM,  
 Reddy R, Yue H, Tang YT;  
 WPI: 2002-034518/04.  
 N-PSDB; AAD23854.  
 Novel human proteases and polynucleotides encoding the proteases,  
 useful for treating, diagnosing or preventing cell proliferative,  
 cardiovascular, autoimmune/inflammatory, neurological and developmental  
 disorders -  
 Claim 1; Page 136-137; 151pp; English.

(infertility, disruption of estrous and menstrual cycle and  
 gynaecomastia); and developmental disorders (renal tubular acidosis,  
 Cushing's syndrome, seizure disorders, congenital glaucoma and cataract).  
 PRS DNA is also in useful in gene therapy. PRS and its immunogenic  
 fragments are useful for screening libraries of compounds in several drug  
 screening assays. The present sequence is human protease PRS-13 protein.  
 SQ Sequence 691 AA;  
 Query Match 17.9%; Score 99; DB 23; Length 691;  
 Best Local Similarity 39.6%; Pred. No. 0.048;  
 Matches 19; Conservative 11; Mismatches 18; Indels 0; Gaps 0;  
 Qy 46 GFDAIYVLAQMVNLQSGIRVISPQPSDIPPGTGVPIVSGGRDND 93  
 127 GADIALLELEPVPVSSHVHTVTLPPASFTFPPGMPGVTCGVDND 174  
 RESULT 22  
 ABB57252  
 ID ABB57252 standard; Protein; 176 AA.  
 AC ABB57252;  
 DT 07-MAR-2002 (first entry)  
 DE Mouse ischaemic condition related protein sequence SEQ ID NO:691.  
 XX  
 KM Mouse; ischaemia; compressive ischaemia; occlusive ischaemia;  
 KM vasoospastic ischaemia; ischaemic condition; ischaemic disease.  
 OS Mus musculus.  
 PN WO200188188-A2.  
 XX 22-NOV-2001.  
 PD 18-MAY-2001; 2001WO-JF04192.  
 PF 18-MAY-2000; 2000JP-0145977.  
 PR 18-MAY-2000; 2000JP-0145977.  
 XX (UNIV-) UNIV NIHON SCHOOL JURIDICAL PERSON.  
 PA Ishikawa K, Asai S, Takahashi Y, Nagata T, Ishii Y;  
 PI WPI: 2002-034733/04.  
 DR N-PSDB; ABI99659.  
 XX Examining the ischemic condition (e.g. occlusive ischemia) by measuring  
 PT expression levels of particular genes defined in the specification or  
 PT by determining the expression profile of a gene group comprising these  
 PT genes -  
 Claim 2; Page 1721-1722; 2690pp; English.

The present invention describes a method for examining ischaemic  
 conditions, comprising measuring the expression levels of particular  
 genes (I) in a test sample or determining the expression profile of a  
 gene group in the sample comprising genes selected from (I). The method  
 is useful for examining the ischaemic condition (e.g. compressive  
 ischaemia, occlusive ischaemia or vasoospastic ischaemia) by measuring  
 CC expression levels of particular genes (ABI99902 to ABI99912, encoding  
 CC the protein sequences of particular genes (ABI99902 to ABI99912) or by determining the  
 CC expression profile of a gene group comprising these genes. The  
 CC expression levels or expression profiles produced by these genes are  
 CC used as an indicator when screening for ischaemic condition-improving  
 CC drugs or therapeutics for ischaemic diseases. ABI99913 and ABI99914  
 CC represent PCR primers for a mouse ischaemic condition related sequence,  
 CC which are used in the exemplification of the present invention.

Query Match 17.4%; Score 96; DB 23; Length 176;





XX PF 25-JUN-1985; 85JP-0138494.  
 XX PR 25-JUN-1985; 85JP-0138494.  
 XX PA (SANKYO ) SANKYO KK.  
 XX DR WPI; 1987-040875/06.  
 XX DR N-PSDB; AAN71122.  
 XX PT Pancreas elastase prodn. used to improve lipoprotein metabolism -  
 PT computes isolating RNA coding elastase, synthesising single and  
 PT double chain cDNA and introducing recombinant into host  
 XX PS Disclosure; Page 481; 18pp; Japanese.  
 XX CC Sequence encoding pancreas elastase may be used to transform an  
 CC expression system. The product is useful in rejuvenating the  
 CC elasticity of the arterial wall, and improving unusual serum lipid  
 CC levels and lipoprotein metabolism.  
 XX SQ Sequence 270 AA;  
 Query Match 17.3%; Score 95.5; DB 8; Length 270;  
 Best Local Similarity 29.1%; Pred. No. 0.038;  
 Matches 25; Conservative 16; Mismatches 34; Indels 11; Gaps 1;  
 QY 4 LTLRNGDQGIHQPSGVKAPGVMPSCMSARQRRPIAQTLSGPDIAIVMLAQVNLQSG 63  
 DB 90 LAVEGPEQVPIINSEELFVHPLNMRSCVAC-----GNDIALIKLSRQAQLGDA 138  
 QY 64 IRVITSLPQPSDIPPPGTVFIVGYR 89  
 DB 139 VQLASLPAPGDIPLPKTPTCYITGWGR 164  
 RESULT 26  
 ID AAB54243 standard; Protein; 270 AA.  
 XX AAU87691;  
 XX AC 21-MAY-2002 (first entry)  
 XX DT 21-MAY-2002 (first entry)  
 XX DE Human pancreatic tumour protein #3.  
 XX KW Human; pancreatic tumour protein; immune response; pancreatic cancer;  
 KW development of cancer; cancer progression; cytostatic.  
 XX OS Homo sapiens.  
 XX PN WO200212331-A2.  
 XX PD 14-FEB-2002.  
 XX PF 06-AUG-2001; 2001WO-US24619.  
 XX PR 07-AUG-2000; 2000US-223130P.  
 PR 30-JAN-2001; 2001US-265447P.  
 PR 15-MAY-2001; 2001US-291201P.  
 XX PA (CORI-) CORIXA CORP.  
 XX PI Pyle RA, Xu J, Kalos MD;  
 XX DR WPI; 2002-241741/29.  
 DR N-PSDB; ABK44206.  
 XX PT Novel polynucleotide encoding pancreatic tumour polypeptides, useful in  
 PT pharmaceutical compositions, e.g. vaccines, for treating pancreatic  
 PT cancer -  
 XX PS Claim 2; Page 162-163; 167pp; English.

XX CC The present invention relates to the isolation of cDNA sequences  
 CC (ABK44061-ABK44209) encoding human pancreatic tumour proteins. The  
 CC polynucleotide sequences encoding human pancreatic tumour proteins are  
 CC useful for stimulating an immune response in a patient and treating  
 CC pancreatic cancer in a patient. A host cell that expresses these  
 CC polynucleotides is useful for determining the presence of cancer in a  
 CC patient. A composition comprising the polynucleotide, its encoded  
 CC protein, or an antibody that binds to the protein may be used in the  
 CC diagnosis, prevention and/or treatment of diseases, particularly  
 CC pancreatic cancer. The sequences of the invention are also useful in  
 CC pharmaceutical compositions, e.g. vaccines, for the diagnosis and  
 CC treatment of pancreatic cancer. Such compositions may be useful for  
 CC inhibiting the development of cancer in a patient, or as markers for  
 CC the progression of cancer. The polynucleotide sequences may also be used  
 CC as probes or primers for nucleic acid hybridisation assays.  
 CC AAU87689-AAU87694 represent human pancreatic tumour proteins.  
 XX SQ Sequence 270 AA;  
 Query Match 17.3%; Score 95.5; DB 23; Length 270;  
 Best Local Similarity 29.1%; Pred. No. 0.038;  
 Matches 25; Conservative 16; Mismatches 34; Indels 11; Gaps 1;  
 QY 4 LTLRNGDQGIHQPSGVKAPGVMPSCMSARQRRPIAQTLSGPDIAIVMLAQVNLQSG 63  
 DB 90 LAVEGPEQVPIINSEELFVHPLNMRSCVAC-----GNDIALIKLSRQAQLGDA 138  
 QY 64 IRVITSLPQPSDIPPPGTVFIVGYR 89  
 DB 139 VQLASLPAPGDIPLPKTPTCYITGWGR 164  
 RESULT 27  
 ID AAB54243 standard; Protein; 273 AA.  
 XX AAB54243;  
 XX AC 09-MAR-2001 (first entry)  
 XX DT 09-MAR-2001 (first entry)  
 XX DE Human pancreatic cancer antigen protein sequence SEQ ID NO:695.  
 XX KW Human; pancreas; pancreatic cancer; pancreatic cancer antigen;  
 KW detection; diagnosis; identification; cytostatic; neuroprotective;  
 KW neotropic; immunomodulatory; relaxant; contraceptive; gynaecological;  
 KW antiinflammatory; cardiant; gene therapy; chromosome mapping;  
 KW linkage analysis; tissue identification; tissue typing; forensic;  
 KW neural; immune system; muscular; reproductive; gastrointestinal;  
 KW pulmonary; cardiovascular; renal; proliferative.  
 XX OS Homo sapiens.  
 XX PN WO200055320-A1.  
 XX PD 21-SEP-2000.  
 XX PF 08-MAR-2000; 2000WO-US05989.  
 XX PR 12-MAR-1999; 99US-0124270.  
 XX PA (HUMA-) HUMAN GENOME SCI INC.  
 XX PI Rosen CA, Ruben SM;  
 XX DR WPI; 2000-579444/54.  
 DR N-PSDB; AAC99008.  
 XX PT New nucleic acid that is a pancreatic cancer antigen for preventing,  
 PT treating, or ameliorating a medical condition, particular pancreatic  
 PT cancer, or for use in assays for diagnosing a pathological condition -  
 XX PS Claim 11; Page 1131-1132; 1379pp; English.

XX AAC98773 to AAC99231 encode the human pancreatic cancer associated  
 CC proteins, called pancreatic cancer antigens, given in AAB54008 to  
 CC AAB54466. The human pancreatic cancer antigens have cytostatic,  
 CC neuroprotective, nootropic, immunomodulatory, relaxant, contraceptive,  
 CC gynaeological, cardiant and antiinflammatory activities, and can be used  
 CC in gene therapy. The polynucleotide and proteins can be used for  
 CC preventing, treating, or ameliorating a medical condition or in assays  
 CC for diagnosing a pathological condition or a susceptibility to one in a  
 CC subject. Binding partners to the proteins and the activity of the  
 CC proteins can be identified. The pancreatic cancer antigens can be used to  
 CC detect, treat or prevent pancreatic disorders, especially cancer.  
 CC Agonists and antagonists to the antigens can be screened for.  
 CC pancreatic cancer antigen polynucleotides can be used to design nucleic  
 CC acid hybridization probes that can be used in chromosome mapping, linkage  
 CC analysis, tissue identification and/or typing and a variety of forensic  
 CC and diagnostic methods. The proteins can be used to generate antibodies  
 CC which are used to purify, detect and target the polypeptides, including  
 CC both in vivo and in vitro diagnostic and therapeutic methods. The  
 CC proteins can be used to treat or prevent neural, immune system, muscular,  
 CC reproductive, gastrointestinal, pulmonary, cardiovascular, renal or  
 CC proliferative disorders. AAC99232 to AAC99240 and AAB54467 represent  
 CC sequences used in the exemplification of the present invention.

SQ Sequence 273 AA;  
 Query Match 17.3%; Score 95.5; DB 21; Length 273;  
 Best Local Similarity 29.1%; Pred. No. 0.038;  
 Matches 25; Conservative 16; Mismatches 34; Indels 11; Gaps 1;

QY 4 LTRNGDGGIHHQPSGVKAVGYPWSCMARQRRIATGSGFDIAIVMLAQWNLQSG 63  
 Db 93 LAKEGPGVPIPIINSEELFVHPLMNRSCVAC-----GNDIALIKLSRSQLGDA 141  
 QY 64 IRVISTPQSPDIPPGTGVIVGYCR 89  
 Db 142 VQLASLPAPAGDILPNKTPCYITWGCR 167

RESULT 28  
 AAY25925  
 ID AAY25925 standard; protein, 244 AA.  
 XX  
 AC AAY25925;  
 XX  
 DT 05-OCT-1999 (first entry)  
 XX  
 DE Human lung mast cell beta-tryptase protein.  
 XX  
 KW Beta-tryptase; mast cell; antiallergic; antiinflammatory; antiasthmatic;  
 KW antiarthritic; inflammatory mediator; treatment; diagnosis; dermatitis;  
 KW allergic; inflammatory disorder; tryptase release; allergic bronchitis;  
 KW conjunctivitis; rhinitis; bronchial asthma; autoimmune disease; human;  
 KW arthritis; multiple sclerosis; psoriasis; inflammatory bowel disease;  
 KW tryptase inhibitor; bifunctional.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO9940073-A2.  
 XX  
 PD 12-AUG-1999.  
 XX  
 PF 04-FEB-1999; 99WO-EP00727.  
 XX  
 PR 06-NOV-1998; 98DE-1051300.  
 PR 06-FEB-1998; 98DE-1004761.  
 XX  
 PA (BYKG ) BYK GULDEN LOMBERG CHEM FAB.  
 PA (PLAC ) MAX PLANCK GES FOERDERUNG WISSENSCHAFTEN.  
 XX  
 PI Baer T, Bergner A, Beume R, Bode W, Bundschuh D,  
 PI Domink A, Goebel K, Huber R, Martin T, Moroder L,  
 PI Pereira PJB, Schaschke N, Sommerhoff C, Stadlwieser J;

PI Thibaut U, Ulrich W;  
 XX  
 DR WPI; 1999-479417/40.  
 XX  
 PT New bifunctional inhibitors of tryptase, used for treating and  
 PT diagnosing allergy and inflammation  
 PS  
 XX Disclosure; Fig 20; 265pp; German.

CC This invention describes novel bifunctional human tryptase (HTR)  
 CC inhibitors (I) of formula K1-L-K2 and their N-oxides (if a heteroaryl,  
 CC heteroarylene, heterocycloalkyl or heterocycloalkylene group is present)  
 CC and salts where K1, K2 = head groups, each including a group Q which  
 CC can interact with carboxylate, b = linker of formula -A5-B5-A3-B3-A1-B1-  
 CC M-B2-A2-A4-B6-A6-, which can adopt a conformation such that the K  
 CC groups are 20-45 Angstrom apart and permits (I) to penetrate into a  
 CC volume of 52 x 32 x 40 Angstrom. A1, A2 = CO, NH, O, S, SO2, SO2NH,  
 CC NH-SO2, CONH, NHCO, OCO, COO or direct bond, A3, A4 = CO, CS, S, S, NH,  
 CC OCO, COO, CONH, NHCO, direct bond or -W-Ra-, W = CO or bond, Ra =  
 CC cyclohexylene, phenylene, oxazolidine-2,3-, 3,4- or 3,5-diy1,  
 CC pyrrolidine-1,2- or 1,3-diy1, piperazine-1,4-diy1, morpholine-2,4- or  
 CC 3,4-diy1, thiamorpholine-2,4- or 3,4-diy1 or piperidine-1,2-, 1,3- or  
 CC 1,4-diy1, A5, A6 = CO, NH, O, S, CONH, NHCO, OCO, COO or direct bond,  
 CC W = -Ar-, -Ar-CR1R2-Ar-, -Ar-Ar-, -Ar-G-Ar-, - (cyclohexylene)-E-  
 CC (cyclohexylene) - (where each ring is optionally substituted by 1-3 A');  
 CC 2,5-dioxopiperazine-3,6-diy1; pyridinediy1 substituted by R5-R7;  
 CC -(CH2)2CONHCH(COR8)(CH2)4-; 5-8c cycloalkylene (optionally substituted  
 CC by 1-3 A'); or a tricyclic group of formula (I)-(IV) (where each benzo  
 CC ring of (I) and (II) is optionally substituted by 1-3 A'), Ar =  
 CC phenylene (optionally substituted by 1-3 A'), R1, R2 = H, A' (optionally  
 CC partially or completely substituted by F) or OH, or CR1R2 = S- or  
 CC 6-membered, optionally substituted cyclic hydrocarbonyl; E, T = CH2, O or  
 CC bond; G = O, S or SO2; R5, R6 = H or A'; R7 = H, A', Ph or pyridyl; R8 =  
 CC OH, NH2, NHA', N(A')2, piperidino or morpholino; K1 = -B7(CO)m-B9-X1,  
 CC -B7(CO)m-B9-Y1 or -B7-(CO)m-B9-B11-X1; K2 = -B8(CO)p-B10-X2,  
 CC -B8(CO)p-B10-Y2 or -B8(CO)p-B10-B12-X2; B1-B6 = bond or 1-4c alkylene;  
 CC B7-B12 = bond or 1-3c alkylene; m, p = 0 or 1; X1, X2 = NH2, SA',  
 CC -C(NH)-NHRD, guanidino, -O-N=C(NH2)2, -S-C(NH)-NH2, -C(NH)-SA',  
 CC -NH-C(NH)-C(NH)-NH2 or -C(NH)-NH2; Rb = H, OH or A';  
 CC A' = 1-4c alkyl; Y1, Y2 = 4-11c heteroaryl or 2-7c heterocycloalkyl,  
 CC containing at least one ring N which can function as proton acceptor  
 CC or donor; Z1, Z2 = 5-12c arylene, 5-12c heteroarylene, 3-8c cycloalkylene  
 CC or 3-8c heterocycloalkylene; arylene, heteroarylene, cycloalkylene,  
 CC 1-heterocycloalkylene, heteroaryl or heterocycloalkyl moieties are  
 CC optionally substituted by 1-3 of OH, halo, NO2, CN, NH2, A', OA', COOA',  
 CC OCOA', COOH or CONH2. The products of the invention have antiallergic,  
 CC antiinflammatory, antiasthmatic and antiarthritic activity. HTR is an  
 CC inflammatory mediator released from mast cells. (I) are used for the  
 CC treatment and diagnosis of allergic and/or inflammatory disorders  
 CC associated with tryptase release, e.g. allergic bronchitis,  
 CC conjunctivitis or rhinitis, bronchial asthma, autoimmune diseases such  
 CC as arthritis, multiple sclerosis, dermatitis and psoriasis and  
 CC inflammatory bowel disease. This sequence represents a human lung mast  
 CC cell beta-tryptase which is used to describe the method of the invention.

SQ Sequence 244 AA;  
 Query Match 16.8%; Score 93; DB 20; Length 244;  
 Best Local Similarity 37.5%; Pred. No. 0.062;  
 Matches 18; Conservative 11; Mismatches 19; Indels 0; Gaps 0;

QY 46 GFDIAIVMLAQWNLQSGIRVISTPQSPDIPPGTGVIVGYCRDND 93  
 Db 89 GADIALLEBPVAVKSSHVHTVTLPPASSETPPGMPGCVITGMDVND 136

RESULT 29  
 AAU12008  
 ID AAU12008 standard; Protein; 245 AA.  
 XX  
 AC AAU12008;  
 XX  
 DT 09-APR-2002 (first entry)

```

XX DE Recombinant human mature beta-II tryptase.
XX KM Human; proteolytic tryptase; protease; recombinant beta-II tryptase;
XX KM enzyme; mutein; mutant.
XX OS Homo sapiens.
XX OS Synthetic.
XX PN WO200198470-A2.
XX PD 27-DEC-2001.
XX PF 20-JUN-2001; 2001WO-US19681.
XX PR 21-JUN-2000; 2000US-0598982.
XX PA (PROM-) PROMEGA CORP.
XX PI Maffei M, Niles AL, Haak-Frendascho M;
XX DR WPI; 2002-114578/15.
XX DR N-PSDB; AAS20766.
XX PT DNA construct for producing enzymatically-inactive proteolytic
XX PT tryptase, comprises DNA sequence encoding proteolytic tryptase having
XX PT an active site mutation -
XX PS Claim 61, Page 80; 126pp; English.
XX CC The present invention relates to recombinant human proteolytic
XX CC tryptases, active site mutants of these tryptases and the methods for
XX CC producing these. The method involves the production of a DNA expression
XX CC construct comprising a promoter operably linked to a secretion signal
XX CC sequence which is operably linked to a DNA sequence encoding a
XX CC proteolytic tryptase with an active site mutation (the construct drives
XX CC expression of a mature proteolytic tryptase that lacks enzymatic activity
XX CC due to the active site mutation, in hosts transformed to contain the
XX CC construct). The method is useful for producing enzymatically-active
XX CC beta-II tryptase. The active site mutants of proteolytic tryptase provide
XX CC a tool to investigate the structural and functional properties of the
XX CC protease and its enzymatic activity, and for modelling studies. The
XX CC enzymatically-active, recombinant proteolytic tryptase produced are
XX CC useful as an antigen to generate anti-human tryptase antibodies
XX CC and in drug screening for compounds which act as tryptase inhibitors,
XX CC antagonists, agonists, etc. The present sequence represents recombinant
XX CC human mature beta-II tryptase.
XX SQ Sequence 245 AA;
XX
XX Query Match 16.8%; Score 93; DB 23; Length 245;
XX Best Local Similarity 37.5%; Pred. No. 0.062;
XX Matches 18; Conservative 11; Mismatches 19; Indels 0; Gaps 0;
XX
XX QY 46 GFDIAIWMALQMVNLQSGIRVLSLPQPSDIPPGTGVFIVGYGRDND 93
XX ID AAVU12013
XX AC AAVU12013 standard; Protein; 245 AA.
XX XX
XX XX 09-APR-2002 (first entry)
XX DE Human beta-II tryptase active site mutant H44A #2.
XX KM Human; proteolytic tryptase; protease; recombinant beta-II tryptase;
XX KM enzyme; mutant; mutein.
XX OS Homo sapiens.

```

```

OS Synthetic.
XX PN WO200198470-A2.
XX PD 27-DEC-2001.
XX PF 20-JUN-2001; 2001WO-US19681.
XX PR 21-JUN-2000; 2000US-0598982.
XX PA (PROM-) PROMEGA CORP.
XX PI Maffei M, Niles AL, Haak-Frendascho M;
XX DR WPI; 2002-114578/15.
XX DR N-PSDB; AAS20779.
XX PT DNA construct for producing enzymatically-inactive proteolytic
XX PT tryptase, comprises DNA sequence encoding proteolytic tryptase having
XX PT an active site mutation -
XX PS Claim 40; Page 95-96; 126pp; English.
XX CC The present invention relates to recombinant human proteolytic
XX CC tryptases, active site mutants of these tryptases and the methods for
XX CC producing these. The method involves the production of a DNA expression
XX CC construct comprising a promoter operably linked to a secretion signal
XX CC sequence which is operably linked to a DNA sequence encoding a
XX CC proteolytic tryptase with an active site mutation (the construct drives
XX CC expression of a mature proteolytic tryptase that lacks enzymatic activity
XX CC due to the active site mutation, in hosts transformed to contain the
XX CC construct). The method is useful for producing enzymatically-active
XX CC beta-II tryptase. The active site mutants of proteolytic tryptase provide
XX CC a tool to investigate the structural and functional properties of the
XX CC protease and its enzymatic activity, and for modelling studies. The
XX CC enzymatically-active, recombinant proteolytic tryptase produced are
XX CC useful as an antigen to generate anti-human tryptase antibodies
XX CC and in drug screening for compounds which act as tryptase inhibitors,
XX CC antagonists, agonists, etc. AAVU2009-AAU1024 represent recombinant
XX CC human beta-II tryptase active site mutants.
XX SQ Sequence 245 AA;
XX
XX Query Match 16.8%; Score 93; DB 23; Length 245;
XX Best Local Similarity 37.5%; Pred. No. 0.062;
XX Matches 18; Conservative 11; Mismatches 19; Indels 0; Gaps 0;
XX
XX QY 46 GFDIAIWMALQMVNLQSGIRVLSLPQPSDIPPGTGVFIVGYGRDND 93
XX ID AAVU12015
XX AC AAVU12015;
XX XX
XX XX 09-APR-2002 (first entry)
XX DE Human beta-II tryptase active site mutant S194A #3.
XX KM Human; proteolytic tryptase; protease; recombinant beta-II tryptase;
XX KM enzyme; mutant; mutein.
XX OS Homo sapiens.
XX OS Synthetic.
XX PN WO200198470-A2.
XX PD 27-DEC-2001.
XX PF 20-JUN-2001; 2001WO-US19681.

```

XX 21-JUN-2000; 2000US-0598982.  
XX (PROM-) PROMEGA CORP.  
XX Maffit M, Niles AL, Haak-Frendscho M;  
PI WPI; 2002-114578/15.  
DR N-PSDB; AAS20781.  
XX  
PT DNA construct for producing enzymatically-inactive proteolytic  
PT trypsinase, comprises DNA sequence encoding proteolytic trypsinase having  
PT an active site mutation -  
XX  
PS Claim 40; Page 100-101; 126pp; English.  
XX  
CC The present invention relates to recombinant human proteolytic  
CC trypsinases, active site mutants of these trypsinases and the methods for  
CC producing these. The method involves the production of a DNA expression  
CC construct comprising a promoter operably linked to a secretion signal  
CC sequence which is operably linked to a DNA sequence encoding a  
CC proteolytic trypsinase with an active site mutation (the construct drives  
CC expression of a mature proteolytic trypsinase that lacks enzymatic activity  
CC due to the active site mutation, in hosts transformed to contain the  
CC construct). The method is useful for producing enzymatically-active  
CC beta-II trypsinase. The active site mutants of proteolytic trypsinase provide  
CC a tool to investigate the structural and functional properties of the  
CC protease and its enzymatic activity, and for modelling studies. The  
CC enzymatically-active, recombinant proteolytic trypsinase produced are  
CC useful as an antigen to generate anti-human trypsinase antibodies  
CC and in drug screening for compounds which act as trypsinase inhibitors,  
CC antagonists, agonists, etc. AAU12009-AAU12024 represent recombinant  
CC human beta-II trypsinase active site mutants.  
XX  
SQ Sequence 245 AA;  
Query Match 16.8%; Score 93; DB 23; Length 245;  
Best Local Similarity 37.5%; Pred. No. 0.062;  
Matches 18; Conservative 11; Mismatches 19; Indels 0; Gaps 0;  
OY 46 GPDIAIWLAAQWNLQSGIRVLSLPQSPDIPPGGVGVIGYGRDND 93  
DB 89 GADIALLEBPVKVSHVHTVTLPPASSTPPGMPGCVTGWGDVND 136  
RESULT 32  
AAU12016  
ID AAU12016 standard; Protein; 245 AA.  
XX  
AC AAU12016;  
XX  
DT 09-APR-2002 (first entry)  
XX  
DE Human beta-II trypsinase active site mutant S194A #4.  
XX  
KW Human; proteolytic trypsinase; protease; recombinant beta-II trypsinase;  
KW enzyme; mutant; mutain.  
XX  
OS Homo sapiens.  
OS Synthetic.  
XX  
PN WO200198470-A2.  
XX  
PD 27-DEC-2001.  
XX  
PF 20-JUN-2001; 2001WO-US19681.  
XX  
PR 21-JUN-2000; 2000US-0598982.  
XX  
PA (PROM-) PROMEGA CORP.  
PI Maffit M, Niles AL, Haak-Frendscho M;  
XX

DR WPI; 2002-114578/15.  
DR N-PSDB; AAS20782.  
XX  
PT DNA construct for producing enzymatically-inactive proteolytic  
PT trypsinase, comprises DNA sequence encoding proteolytic trypsinase having  
PT an active site mutation -  
XX  
PS Claim 40; Page 102-103; 126pp; English.  
XX  
CC The present invention relates to recombinant human proteolytic  
CC trypsinases, active site mutants of these trypsinases and the methods for  
CC producing these. The method involves the production of a DNA expression  
CC construct comprising a promoter operably linked to a secretion signal  
CC sequence which is operably linked to a DNA sequence encoding a  
CC proteolytic trypsinase with an active site mutation (the construct drives  
CC expression of a mature proteolytic trypsinase that lacks enzymatic activity  
CC due to the active site mutation, in hosts transformed to contain the  
CC construct). The method is useful for producing enzymatically-active  
CC beta-II trypsinase. The active site mutants of proteolytic trypsinase provide  
CC a tool to investigate the structural and functional properties of the  
CC protease and its enzymatic activity, and for modelling studies. The  
CC enzymatically-active, recombinant proteolytic trypsinase produced are  
CC useful as an antigen to generate anti-human trypsinase antibodies  
CC and in drug screening for compounds which act as trypsinase inhibitors,  
CC antagonists, agonists, etc. AAU12009-AAU12024 represent recombinant  
CC human beta-II trypsinase active site mutants.  
XX  
SQ Sequence 245 AA;  
Query Match 16.8%; Score 93; DB 23; Length 245;  
Best Local Similarity 37.5%; Pred. No. 0.062;  
Matches 18; Conservative 11; Mismatches 19; Indels 0; Gaps 0;  
OY 46 GPDIAIWLAAQWNLQSGIRVLSLPQSPDIPPGGVGVIGYGRDND 93  
DB 89 GADIALLEBPVKVSHVHTVTLPPASSTPPGMPGCVTGWGDVND 136  
RESULT 33  
AAU55011  
ID AAU55011 standard; Protein; 249 AA.  
XX  
AC AAU55011;  
XX  
DT 18-FEB-2000 (first entry)  
XX  
DE Human beta-tryptase protein sequence.  
XX  
KW Beta-tryptase; human; DNA expression construct; protein production;  
KW combinatorial library screening; X ray crystallography; antigen;  
KW antibody generation.  
XX  
OS Homo sapiens.  
XX  
PN WO9960139-A1.  
XX  
PD 25-NOV-1999.  
XX  
PF 29-OCT-1998; 98WO-US22994.  
XX  
PR 15-MAY-1998; 98US-0079970.  
XX  
PA (PROM-) PROMEGA CORP.  
PI Maffit MA, Niles AL, Haak-Frendscho M;  
XX  
XX WPI; 2000-053300/04.  
DR N-PSDB; AAZ40175.  
XX  
PT New DNA expression construct for production of enzymatically active  
PT recombinant human beta-tryptase -  
PS Disclosure; Page 43-44; 50pp; English.

XX This sequence is the human beta-tryptase. The invention relates to a  
CC DNA expression construct comprising (5' to 3') a promoter linked to a  
CC signal sequence which is linked to a sequence encoding human  
CC beta-tryptase. The DNA construct is useful for transforming host cells to  
CC express, post translationally process and secrete enzymatically active  
CC human tryptase. The method is useful for the production of large amounts  
CC of tryptase with defined specifications. The transformant is useful for  
CC pharmacological studies, combinatorial library screens and X ray  
CC crystallographic studies. The tryptase produced allows for the  
CC development of tryptase agonists and/or antagonists, is useful as an  
CC antigen to generate antihuman tryptase antibodies in various animals,  
CC can be used in screening for compounds which act as tryptase inhibitors,  
CC antagonists, agonists etc. and to assay for the presence of tryptase in  
CC biological or other solutions. Tryptase inhibitors, antagonists, agonists  
CC etc. may be useful as therapeutics. The tryptase does not require any  
CC post-expression or post-purification modifications or manipulations to  
CC initiate tryptase activity and it has enzymatic activity which compares  
CC favourably with cadaveric tryptase. The availability of enzymatically  
CC active tryptase facilitates the large scale screening of combinatorial  
CC libraries for specific tryptase inhibitors as potential therapeutics and  
CC advances the understanding of the biological significance of tryptase in  
CC mast cell mediated diseases. The tryptase can be used to detect low  
CC levels of tryptase.

SQ Sequence 249 AA;

Query Match 16.8%; Score 93; DB 21; Length 249;

Best Local Similarity 37.5%; Pred. No. 0.064;

Matches 18; Conservative 11; Mismatches 19; Indels 0; Gaps 0;

QY 46 GFDIAIVMLAQMVNLQSGIRVLSLPQSDIPPPTGTVFVIGYGRDND 93

Dd 93 GADIALLELEPVKVSSHVHTVTLPPASETPPPGMPGVGTGMDVND 140

RESULT 34  
AAU12007  
ID AAU12007 standard; Protein: 249 AA.

AC AAU12007;

DT 09-APR-2002 (first entry)

XX Recombinant human beta-II tryptase.

XX Human; proteolytic tryptase; protease; recombinant beta-II tryptase;

KW enzyme; mutant; mutant.

XX Homo sapiens.

OS Synthetic.

XX W0200198470-A2.

XX 27-DEC-2001.

XX 20-JUN-2001; 2001WO-US19681.

XX 21-JUN-2000; 2000US-0598982.

XX (PROM-) PROMEGA CORP.

XX Maffitt M, Niles AL, Haak-Frendescho M;

XX WPI; 2002-114578/15.

XX N-PSDB; AAS20765.

PT DNA construct for producing enzymatically-inactive proteolytic  
PT tryptase, comprises DNA sequence encoding proteolytic tryptase having  
PT an active site mutation -

XX Claim 60; Page 77-78; 126pp; English.

CC The present invention relates to recombinant human proteolytic  
CC tryptases, active site mutants of these tryptases and the methods for  
CC producing these. The method involves the production of a DNA expression  
CC construct comprising a promoter operably linked to a secretion signal  
CC sequence which is operably linked to a DNA sequence encoding a  
CC proteolytic tryptase with an active site mutation (the construct drives  
CC expression of a mature proteolytic tryptase that lacks enzymatic activity  
CC due to the active site mutation, in hosts transformed to contain the  
CC construct). The method is useful for producing enzymatically-active  
CC beta-II tryptase. The active site mutants of proteolytic tryptase provide  
CC a tool to investigate the structural and functional properties of the  
CC enzyme and its enzymatic activity, and for modelling studies. The  
CC enzymatically-active, recombinant proteolytic tryptase produced are  
CC useful as an antigen to generate anti-human tryptase antibodies  
CC and in drug screening for compounds which act as tryptase inhibitors,  
CC antagonists, agonists, etc. The present sequence represents recombinant  
CC human beta-II tryptase.

SQ Sequence 249 AA;

Query Match 16.8%; Score 93; DB 23; Length 249;

Best Local Similarity 37.5%; Pred. No. 0.064;

Matches 18; Conservative 11; Mismatches 19; Indels 0; Gaps 0;

QY 46 GFDIAIVMLAQMVNLQSGIRVLSLPQSDIPPPTGTVFVIGYGRDND 93

Dd 93 GADIALLELEPVKVSSHVHTVTLPPASETPPPGMPGVGTGMDVND 140

RESULT 35  
AAU12009  
ID AAU12009 standard; Protein: 249 AA.

AC AAU12009;

DT 09-APR-2002 (first entry)

XX Human beta-II tryptase active site mutant H44A #1.

XX Human; proteolytic tryptase; protease; recombinant beta-II tryptase;

KW enzyme; mutant; mutant.

XX Homo sapiens.

OS Synthetic.

XX W0200198470-A2.

XX 27-DEC-2001.

XX 20-JUN-2001; 2001WO-US19681.

XX 21-JUN-2000; 2000US-0598982.

XX (PROM-) PROMEGA CORP.

XX Maffitt M, Niles AL, Haak-Frendescho M;

XX WPI; 2002-114578/15.

XX N-PSDB; AAS20775.

PT DNA construct for producing enzymatically-inactive proteolytic  
PT tryptase, comprises DNA sequence encoding proteolytic tryptase having  
PT an active site mutation -

XX Claim 8; Page 84-85; 126pp; English.

CC The present invention relates to recombinant human proteolytic  
CC tryptases, active site mutants of these tryptases and the methods for  
CC producing these. The method involves the production of a DNA expression  
CC construct comprising a promoter operably linked to a secretion signal  
CC sequence which is operably linked to a DNA sequence encoding a  
CC proteolytic tryptase with an active site mutation (the construct drives  
CC expression of a mature proteolytic tryptase that lacks enzymatic activity

CC due to the active site mutation, in hosts transformed to contain the  
CC construct). The method is useful for producing enzymatically-active  
CC beta-II tryptase. The active site mutants of proteolytic tryptase provide  
CC a tool to investigate the structural and functional properties of the  
CC protease and its enzymatic activity, and for modelling studies. The  
CC enzymatically-active, recombinant proteolytic tryptase produced are  
CC useful as an antigen to generate anti-human tryptase antibodies,  
CC and in drug screening for compounds which act as tryptase inhibitors,  
CC antagonists, agonists, etc. AAU12009-AAU12024 represent recombinant  
CC human beta-II tryptase active site mutants.

XX  
SQ Sequence 249 AA;

Query Match 16.8%; Score 93; DB 23; Length 249;  
Best Local Similarity 37.5%; Pred. No. 0.064;  
Matches 18; Conservative 11; Mismatches 19; Indels 0; Gaps 0;

Qy 46 GFDIAIWLAAQMVNLQSGIRVISTLPQPSDIPPGTGIVFGGRDND 93  
Db 93 GADIALLEBPVKVSHVHTVTLPPASSTFPFGMPCMWVGMDVND 140

RESULT 36  
AAU12011  
ID AAU12011 standard; Protein; 249 AA.  
XX  
AC AAU12011;  
XX  
DT 09-APR-2002 (first entry)  
XX  
DE Human beta-II tryptase active site mutant S194A #1.  
XX  
KW Human; proteolytic tryptase; protease; recombinant beta-II tryptase;  
KW enzyme; mutant; mutein.  
XX  
OS Homo sapiens.  
OS Synthetic.  
OS  
PN WO200198470-A2.  
XX  
PD 27-DEC-2001.  
XX  
PF 20-JUN-2001; 2001WO-US19681.  
XX  
PR 21-JUN-2000; 2000US-0598982.  
XX  
PA (PROM-) PROMEGA CORP.  
XX  
PI Maffit M, Niles AL, Haak-Frendescho M;  
XX  
DR WPI; 2002-114578/15.  
DR N-PSDB; AAS20777.  
XX  
PT DNA construct for producing enzymatically-inactive proteolytic  
PT tryptase, comprises DNA sequence encoding proteolytic tryptase having  
PT an active site mutation -  
XX  
PS Claim 8; Page 90-91; 126pp; English.

CC The present invention relates to recombinant human proteolytic  
CC tryptases, active site mutants of these tryptases and the methods for  
CC producing these. The method involves the production of a DNA expression  
CC construct comprising a promoter operably linked to a secretion signal  
CC sequence which is operably linked to a DNA sequence encoding a  
CC proteolytic tryptase with an active site mutation (the construct drives  
CC expression of a mature proteolytic tryptase that lacks enzymatic activity  
CC due to the active site mutation, in hosts transformed to contain the  
CC construct). The method is useful for producing enzymatically-active  
CC beta-II tryptase. The active site mutants of proteolytic tryptase provide  
CC a tool to investigate the structural and functional properties of the  
CC protease and its enzymatic activity, and for modelling studies. The  
CC enzymatically-active, recombinant proteolytic tryptase produced are  
CC useful as an antigen to generate anti-human tryptase antibodies,  
CC and in drug screening for compounds which act as tryptase inhibitors,  
CC antagonists, agonists, etc. AAU12009-AAU12024 represent recombinant  
CC human beta-II tryptase active site mutants.

CC and in drug screening for compounds which act as tryptase inhibitors,  
CC antagonists, agonists, etc. AAU12009-AAU12024 represent recombinant  
CC human beta-II tryptase active site mutants.

XX  
SQ Sequence 249 AA;

Query Match 16.8%; Score 93; DB 23; Length 249;  
Best Local Similarity 37.5%; Pred. No. 0.064;  
Matches 18; Conservative 11; Mismatches 19; Indels 0; Gaps 0;

Qy 46 GFDIAIWLAAQMVNLQSGIRVISTLPQPSDIPPGTGIVFGGRDND 93  
Db 93 GADIALLEBPVKVSHVHTVTLPPASSTFPFGMPCMWVGMDVND 140

RESULT 37  
AAU12012  
ID AAU12012 standard; Protein; 249 AA.  
XX  
AC AAU12012;  
XX  
DT 09-APR-2002 (first entry)  
XX  
DE Human beta-II tryptase active site mutant S194A #2.  
XX  
KW Human; proteolytic tryptase; protease; recombinant beta-II tryptase;  
KW enzyme; mutant; mutein.  
XX  
OS Homo sapiens.  
OS Synthetic.  
OS  
PN WO200198470-A2.  
XX  
PD 27-DEC-2001.  
XX  
PF 20-JUN-2001; 2001WO-US19681.  
XX  
PR 21-JUN-2000; 2000US-0598982.  
XX  
PA (PROM-) PROMEGA CORP.  
XX  
PI Maffit M, Niles AL, Haak-Frendescho M;  
XX  
DR WPI; 2002-114578/15.  
DR N-PSDB; AAS20778.  
XX  
PT DNA construct for producing enzymatically-inactive proteolytic  
PT tryptase, comprises DNA sequence encoding proteolytic tryptase having  
PT an active site mutation -  
XX  
PS Claim 8; Page 93; 126pp; English.

CC The present invention relates to recombinant human proteolytic  
CC tryptases, active site mutants of these tryptases and the methods for  
CC producing these. The method involves the production of a DNA expression  
CC construct comprising a promoter operably linked to a secretion signal  
CC sequence which is operably linked to a DNA sequence encoding a  
CC proteolytic tryptase with an active site mutation (the construct drives  
CC expression of a mature proteolytic tryptase that lacks enzymatic activity  
CC due to the active site mutation, in hosts transformed to contain the  
CC construct). The method is useful for producing enzymatically-active  
CC beta-II tryptase. The active site mutants of proteolytic tryptase provide  
CC a tool to investigate the structural and functional properties of the  
CC protease and its enzymatic activity, and for modelling studies. The  
CC enzymatically-active, recombinant proteolytic tryptase produced are  
CC useful as an antigen to generate anti-human tryptase antibodies,  
CC and in drug screening for compounds which act as tryptase inhibitors,  
CC antagonists, agonists, etc. AAU12009-AAU12024 represent recombinant  
CC human beta-II tryptase active site mutants.

XX  
SQ Sequence 249 AA;

Query Match 16.8%; Score 93; DB 23; Length 249;



```

ID AAU84360 standard; Protein; 267 AA.
XX
AC AAU84360;
XX
DT 08-MAY-2002 (first entry)
XX
DE Protein TPS2 differentially expressed in breast cancer tissue.
XX
KM Human; diagnosis of breast cancer; endometrial cancer; breast tumour;
XX MLI; mitotic activity index; cytostatic.
OS Homo sapiens.
XX
PN WO200210436-A2.
XX
PD 07-FEB-2002.
XX
PF 27-JUL-2001; 2001WO-US23642.
XX
PR 28-JUL-2000; 2000US-222093P.
XX
PA (BGM ) BRIGHAM & WOMENS HOSPITAL INC.
XX (BAK/) BAK J.
XX
PI Baak J, Mutter GL;
XX
XX WPI; 2002-180084/23.
XX DR N-PSDB; ABK35580.
XX
PT Diagnosing breast cancer comprises determining expression of nucleic
PT acid molecules or expression products that are differentially expressed
PT in normal and malignant tissue -
XX
PS Claim 37; Page 217-218; 219pp; English.
XX
XX The present invention relates to a method for diagnosing breast cancer
CC in a subject suspected of having endometrial cancer. The method
CC comprises determining the expression of a set of human genes or
CC expression products in an endometrial sample suspected of being
CC cancerous. The human genes of the invention are differentially
CC expressed in breast tumours characterised as high or low MLI (mitotic
CC activity index). These sets of genes can be used to discriminate between
CC high and low MLI breast tumours. The invention also provides DNA and
CC protein microarrays for analysing the expression of the human genes and
CC their protein products. The methods and arrays are useful for the
CC diagnosis and prognosis of endometrial cancer, selecting and monitoring
CC treatment regimes, and identification of compounds useful for the
CC treatment of endometrial cancer. AAU84311-AAU84361 represent the human
CC proteins of the invention that are differentially expressed in breast
CC cancer tissue.
XX
SQ Sequence 267 AA;
XX
Query Match 16.8%; Score 93; DB 23; Length 267;
Best Local Similarity 37.5%; Pred. No. 0.069;
Matches 18; Conservative 11; Mismatches 19; Indels 0; Gaps, 0;
OY 46 GPDIAIWLACQVNLQSGRIVSLPQPSDIPRGVFIYGYGRDND 93
Db 111 GADIALLEBEPVKVSHVHTVTLPPASSTFPGMPCWVWGMDVDND 158

```

Search completed: April 9, 2003, 12:55:27  
 Job time : 90.623 secs



GenCore version 5.1.4.p5.4578  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - protein search, using frame\_n2p model

Run on: April 10, 2003, 07:47:28 ; Search time 87 Seconds

(without alignments)  
4524.395 Million cell updates/sec

Title: US-10-020-441-1

Perfect score: 477

Sequence: 1 ggtcaaccgcgtgatagc.....ttgtacgtccatatttga 1477

Scoring table:

OLIGO

Xgapop 60.0 , Xgapext 60.0

Ygapop 60.0 , Ygapext 60.0

Fgapop 6.0 , Fgapext 7.0

Delop 6.0 , Delext 7.0

Searched: 908470 seqs, 133250620 residues

Word size: 1

Total number of hits satisfying chosen parameters: 1687582

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Listing first 1000 summaries

Command line parameters:

-MODE=frame\_n2p\_model -DEV=xlp

-Q=/cgn2\_1/USPRO/spool/US10020441/runat\_10042003\_074719\_25890/app-query.fasta.1.1671

-DB=Geneseq\_101002 -QFMT=fastan -SUFFIX=oligo.rag -MINMATCH=0.1 -LOOPCL=0

-LIST=1000 -UNITS=bits -START=1 -END=1 -MATRIX=oligo -TRANS=human40.cdi

-OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000

-USRR=US10020441.ecgn.1.114.0runat\_10042003\_074719\_25890 -NCPU=6 -ICPU=3

-NO\_XLPHY -NO\_MMAP -LARGEQUERT -NBS\_SCORES=0 -WAIT -LONGLOG -DEV.TIMEOUT=120

-WARN\_TIMEOUT=30 -THREDS=1 -XGAPOP=60 -XGAPEXT=60 -Fgapop=6 -Fgapext=7

-YGAPOP=60 -YGAPEXT=60 -DELOP=6 -DELEX=7

Database :

1: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.\*

2: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.\*

3: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.\*

4: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.\*

5: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.\*

6: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.\*

7: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.\*

8: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.\*

9: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.\*

10: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.\*

11: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.\*

12: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.\*

13: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.\*

14: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.\*

15: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.\*

16: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.\*

17: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.\*

18: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.\*

19: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.\*

20: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.\*

21: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.\*

22: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.\*

23: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2002.DAT.\*

## SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	106	22.2	106	21	AAV67587	S. mansoni cercari
2	16	3.4	16	21	AAV67588	S. mansoni cercari
3	10	2.1	10	16	AAW21482	Schistosoma elasta
4	10	2.1	34	22	AAO09392	Human polypeptide
5	9	1.9	11	16	AAW21484	Schistosoma elasta
6	9	1.9	40	22	AAO02434	Human polypeptide
7	9	1.9	44	21	AAV64770	Human 5' EST relat
8	9	1.9	55	22	AAO07568	Human polypeptide
9	9	1.9	56	23	ABP00091	Human OREF protein
10	9	1.9	922	22	ABG34997	Novel human dialno
11	8	1.7	28	22	AAO05400	Human polypeptide
12	8	1.7	29	22	AAO04534	Human polypeptide
13	8	1.7	36	21	AAO58999	Arabidopsis thalia
14	8	1.7	47	21	AAO44943	Human secreted pro
15	8	1.7	53	23	ABP32259	Human OREF1232 prot
16	8	1.7	61	22	AAO13377	Human gene 3 encod
17	8	1.7	61	23	ABG64084	Human albumin fusl
18	8	1.7	67	21	AAO58997	Arabidopsis thalia
19	8	1.7	79	22	AAO00781	Human polypeptide
20	8	1.7	82	22	AAO08984	Human polypeptide
21	8	1.7	83	22	AAO84230	Human immune/haema
22	8	1.7	84	21	AAO33279	Human OREF ORF3043
23	8	1.7	90	23	ABP33051	Human ORF2024 prot
24	8	1.7	92	22	AAO02862	Human polypeptide
25	8	1.7	102	21	AAO1641	Human OREF ORF1405
26	8	1.7	103	22	AAO47000	Protonbacterium
27	8	1.7	124	22	AAO04304	Human polypeptide
28	8	1.7	146	22	AAO01446	Human polypeptide
29	8	1.7	176	22	ABG57252	Mouse ischaemic co
30	8	1.7	207	23	ABG66758	Human novel polype
31	8	1.7	213	22	AAO54490	Protonbacterium
32	8	1.7	222	15	AAO53058	Human mature cathe
33	8	1.7	223	17	AAO01165	Serine protease Pf
34	8	1.7	223	22	ABG55927	Drosophila melanog
35	8	1.7	224	22	ABG50561	Flea setine protea
36	8	1.7	226	17	AAO09318	Human wild type ch
37	8	1.7	226	17	AAO09319	Human chymase clon
38	8	1.7	226	17	AAO09320	Human chymase clon
39	8	1.7	227	19	AAO64263	Rat mast cell prot
40	8	1.7	247	20	AAW3878	Human chymase prot
41	8	1.7	265	22	ABG60999	Drosophila melanog
42	8	1.7	278	22	AAO50962	C glutamicum prote
43	8	1.7	278	22	AAO50962	Corynebacterium gl
44	8	1.7	299	10	AAO94148	Signal peptide-pro
45	8	1.7	316	22	ABG55933	Drosophila melanog
46	8	1.7	334	22	ABG55858	Drosophila melanog
47	8	1.7	367	22	ABG17105	Drosophila melanog
48	8	1.7	390	22	ABG57834	Drosophila melanog
49	8	1.7	393	22	ABG58802	Drosophila melanog
50	8	1.7	393	22	ABG58812	Drosophila melanog
51	8	1.7	395	22	ABG52640	Escherichia melano
52	8	1.7	408	22	ABG17152	Drosophila melanog
53	8	1.7	431	10	AAO94764	Non-glycosylated p
54	8	1.7	733	23	ABG19955	LDL receptor bindi
55	8	1.7	829	23	ABO04817	Rat glutamate tran
56	8	1.7	860	23	ABO07689	C. trachomatis B s
57	8	1.7	1013	20	AAV16737	Drosophila melanog
58	8	1.7	1379	22	ABG58860	Conserved serine p
59	8	1.7	9	17	AAO01226	HLA A0201 binding
60	8	1.7	9	21	AAO32274	Human TADG-12 immu
61	8	1.7	9	22	AAO68956	Human stratum corn
62	8	1.7	9	22	AAO08289	HLA binding TADG-1
63	8	1.7	9	22	AAO02291	Flea highly conser
64	8	1.7	9	22	AAO50560	Rheumatoid arthrit
65	8	1.7	11	20	AAO42050	DPI tryptic digest
66	8	1.7	11	22	AAO28497	Depression-Associa
67	8	1.7	11	22	AAO26143	Peptide fragment #
68	8	1.7	12	21	AAV72111	

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.













945	7	1.5	272	23	AA082741	Amino acid sequenc
946	7	1.5	273	13	AA025397	Recombinant scu-PA
947	7	1.5	273	13	AA025398	Recombinant scu-PA
948	7	1.5	273	19	AA064238	Human mast cell tr
949	7	1.5	273	19	AA064238	Murine mMcP-7 zymo
950	7	1.5	273	19	AA064235	Rat homologue of m
951	7	1.5	273	19	AA063174	Human mast cell tr
952	7	1.5	273	21	AA054243	Human pancreatic c
953	7	1.5	273	21	AA011696	Mouse serine prote
954	7	1.5	273	22	AA061377	Drosophila melanog
955	7	1.5	273	22	AA093816	Human polypeptide
956	7	1.5	273	22	AA015594	Gene #9 human secr
957	7	1.5	273	22	AA015585	Human secreted pro
958	7	1.5	274	13	AA025359	Recombinant scu-PA
959	7	1.5	274	13	AA025400	Recombinant scu-PA
960	7	1.5	274	13	AA025403	Recombinant scu-PA
961	7	1.5	274	13	AA025404	Recombinant scu-PA
962	7	1.5	274	19	AA064240	Human mast cell tr
963	7	1.5	274	19	AA064234	Rat homologue of m
964	7	1.5	274	19	AA063175	Human mast cell tr
965	7	1.5	274	22	AA062403	Drosophila melanog
966	7	1.5	274	22	AA024196	Human K2S heterolo
967	7	1.5	275	19	AA064237	Human mast cell tr
968	7	1.5	275	19	AA063173	Human mast cell tr
969	7	1.5	275	21	AA021311	Human neuropilin.
970	7	1.5	275	21	AA011714	Human serine prote
971	7	1.5	275	21	AA01826	Pig lung protease.
972	7	1.5	275	22	AA008589	Human NCV10 protei
973	7	1.5	276	19	AA064242	Murine mast cell p
974	7	1.5	276	19	AA063172	Mouse mast cell pr
975	7	1.5	276	21	AA011713	Mouse serine prote
976	7	1.5	276	22	AA064035	Drosophila melanog
977	7	1.5	276	22	AA064108	Drosophila melanog
978	7	1.5	276	22	AA021759	Novel human diagno
979	7	1.5	276	23	AA016546	Human LKW-urokinas
980	7	1.5	276	23	AA005437	Mouse mast cell pr
981	7	1.5	277	21	AA021300	Human KIK-14 prote
982	7	1.5	277	21	AA023912	Porcine virus acti
983	7	1.5	277	22	AA065958	Drosophila melanog
984	7	1.5	278	21	AA058142	Lung cancer associ
985	7	1.5	278	22	AA010063	Mast cell protease
986	7	1.5	279	17	AA094526	Korean Viper Salmo
987	7	1.5	279	22	AA065085	Drosophila melanog
988	7	1.5	279	22	AA020513	Novel human diagno
989	7	1.5	279	23	AA077011	Human eosinophil s
990	7	1.5	280	15	AA047381	N-terminally modif
991	7	1.5	280	15	AA047582	N-terminally modif
992	7	1.5	280	21	AA042485	Human ORFX ORF2249
993	7	1.5	280	22	AA064195	Drosophila melanog
994	7	1.5	280	23	AA061010	Novel human protei
995	7	1.5	281	13	AA025401	Recombinant scu-PA
996	7	1.5	281	13	AA025402	Recombinant scu-PA
997	7	1.5	281	20	AA042440	CASB12 polypeptide
998	7	1.5	281	23	AA018303	Lumbric rubellus
999	7	1.5	282	20	AA042439	CASB12 amino acid
1000	7	1.5	282	21	AA011712	Human serine prote

## ALIGNMENTS

RESULT 1  
AA067587  
ID AAY67587 standard; Protein: 106 AA.

XX AAY67587;

XX 19-JUN-2000 (first entry)

DE S. mansoni cercarial elastase protein fragment.

XX Schistosoma parasite; fusion protein; cercarial elastase; vaccine;

KW antibody response; schistosomicide.

OS Schistosoma mansoni.  
XX  
XX EP992582-A2.  
XX  
XX 12-APR-2000.  
XX  
XX 05-OCT-1999; 99EP-0307832.  
XX  
XX 07-OCT-1998; 98GB-0021821.  
XX  
XX (UYWA-) UNIV WALES BANGOR.  
XX  
XX Doenhoff M, Sayers J;  
XX  
XX WPI; 2000-259136/23.  
XX  
XX N-PSDB: AA290608.  
XX  
XX New vaccine for treatment of Schistosoma infections contains a  
XX recombinant fusion protein comprising cercarial elastase sequence fused  
XX to bacterial, phage or viral protein -  
XX  
XX Claim 3; Page 17; 26pp; English.

CC The invention provides a vaccine comprising a recombinant fusion protein  
CC capable of eliciting immunity against Schistosoma parasites. The fusion  
CC protein comprises the 27 or 28 kDa cercarial elastase sequence of  
CC S. mansoni or an active fragment, homologue or variant, fused to a  
CC bacterial, phage or viral protein. The vaccine containing the fusion  
CC protein has been found to induce a significantly increased antibody  
CC response against schistosoma infections, compared to the use of S.  
CC mansoni cercarial elastase in its native form. The present sequence  
CC represents the S. mansoni cercarial elastase protein fragment encoded by  
CC exon 2 of the cercarial elastase gene, used in the fusion protein of the  
CC invention.  
XX  
XX

SO Sequence 106 AA;

Alignment Scores:

Pred. No.: 4.15e-92 Length: 106  
Score: 106.00 Matches: 106  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 22.22% Indels: 0  
DB: 21 Gaps: 0

US-10-020-441-1 (1-1477) x AAY67587 (1-106)

QY 310 GTTCATTTCTCACACTGAGGAATGGCGACCAAGCATTCATCCACCGCTGGA 369

DB 1 ValSerPheLeuThrLeuAlaArgAsnGlyaspGlnGnglyIleHisHisGlnProSerGly 20

QY 370 GTTAAAGTGGCAGCAGATACATGCGCTTGTATGTCGCGACGACAGAGACCAATC 429

DB 21 ValIysValAlaProGlyTyrMetProSerGlySmetSerAlaArgGlnAlaArgProIle 40

QY 430 GCACAGACACTCAGTGGATTCGATATTCATTTGAATGCTGCGCTCAATAGTCAACTA 489

DB 41 AlaGlnThrLeuSerGlyPheAspIleAlaIleValMetLeuValGlnMetValAsnLeu 60

QY 490 CAGAGTGAATCAGAGTATGATGCTGCGACGACCATGATCCGCGACCTGGAAT 549

DB 61 GlnSerGlyIleArgValIleSerLeuProGlnProSerAspIleProProGlyThr 80

QY 550 GGTGTTTCATTTGTTGTTATGAAGGATGATAACGACCGTGAATCCGTCACGTAAGAT 609

DB 81 GlyValPheIleValAlaGlyTyrGlyArgAspAspAsnAspArgAspProSerArgLysAsn 100

QY 610 GGTGCAATATGAGAGAA 627

DB 101 GlyGlyIleLeuLysLys 106

RESULT 2  
AAY67588



KW competitive inhibitor; feedback regulator; synthesis; gastrin precursor  
KW charge; polarity; farnesyl synthetase; plasminogen activator inhibitor 1  
KM hydroxymethylglutaryl coenzyme A reductase; glucagon precursor; rhesus;  
KW gonadoliberin precursor; plasminogen activator inhibitor 2; protein; Alzheim  
KW Alzheimer amyloid A4; corticotropin releasing factor binding protein; apoli  
KW apolipoprotein E; herpes virus 1 glycoprotein B; HSV1; human; OMVVS;  
KW herpes virus 2 glycoprotein B; HSV2; collagenase; apolipoprotein A;  
KW Treponema pallidum membrane protein; TMPA; islet amyloid polypeptide;  
KW fibroblast MMP1; schistosoma elastase precursor; schistosomini;  
KW hepatitis delta antigen; rev protein; HIV; VILV; angiotensinogen.  
XX Schistosome sp.  
XX WO9519568-A1.  
PN 20-JUL-1995.  
XX  
PD 12-JAN-1995; 95WO-US00575.  
PF 12-JAN-1995; 95WO-US00575.  
XX  
XX 14-JAN-1994; 94US-0182248.  
PR  
XX (RATH/) RATH M.  
PA  
XX Rath M;  
PI  
XX WPI; 1995-263953/34.  
DR  
XX Identifying signal oligopeptide(s) in protein sequence(s) - shown as  
PT regions of max. hydrophilicity, used in modulating communication  
between protein(s)  
XX  
XX Claim 5; Page 69; 88pp; English.

The sequences given in AAM21201-560 represent hydrophilic signal oligo-peptides. These signal oligopeptides are localised on the surface of the protein and are represented by the hydrophilicity maxima of the protein. These peptides are enriched in charged amino acids arranged with neutral spacer amino acids. The specific signal character of these oligopeptides is determined by a characteristic combination of conformation and charge within the signal sequence. CC These oligopeptides may be used as vaccines in the treatment of human disease, as competitive inhibitors to prevent or reduce the metabolic action or interaction of a selected protein by blocking its specific signal sequences, or as therapeutic agents to function as feedback regulators to reduce synthesis rate of a selected protein. CC These peptides may be modified by omitting one or more amino acids at the N- and/or C-terminal, by substituting one or more amino acids without consideration of charge and polarity, by substituting one or more amino acids with amino acid residues with similar charge and/or polarity, by omitting one or more amino acids or a combination of these.

XX  
XX Sequence 10 AA;

Alignment Scores:

Pred. No.:	Length:
Score: 1.72	Matches: 10
Percent Similarity: 10.00	Conservative: 10
Best Local Similarity: 100.00%	Mismatches: 0
Query Match: 2.10%	Indels: 0
DB: 16	Gaps: 0

US-10-020-441-1 (1-1477) x AAM21482 (1-10)

OY 328 AGCAATGGCGACCAACAAGCATCATCAC 357  
|||||  
Db 1 Argasnclyaspnglngnlglyldehshts 10

RESULT 4  
AAO09392  
ID AAO09392 standard; Protein; 34 AA.  
XX  
XX AAO09392;

DT 06-NOV-2001 (first entry)  
XX  
DE Human polypeptide SEQ ID NO 23284.  
XX  
KW Human; cytokine; cell proliferation; cell differentiation; gene therapy;  
KW vaccine; peptide therapy; stem cell growth factor; haematopoiesis;  
KW tissue growth factor; immunomodulatory; cancer; leukaemia;  
KW nervous system disorders; arthritis; inflammation.  
XX  
OS Homo sapiens.  
XX  
PN MO200164835-A2.  
XX  
PD 07-SEP-2001.  
XX  
PF 26-FEB-2001; 2001WO-US04927.  
XX  
PR 28-FEB-2000; 2000US-0515126.  
PR 18-MAY-2000; 2000US-0577409.  
XX  
PA (HYSE-) HYSEO INC.  
XX  
PI Tang YT, Liu C, Drmanac RT;  
XX  
DR WPI: 2001-514838/56.  
DR N-PSDB; AA189323.  
XX  
PT Isolated nucleic acids and polypeptides, useful for preventing  
PT diagnosing and treating e.g. leukaemia, inflammation and immune  
PT disorders -  
XX  
PS Claim 20; SEQ ID NO 23284; 1399pp + Sequence Listing; English.  
XX  
CC The invention relates to human polynucleotides (AA179941-AA193841) and  
CC the encoded proteins (AA000010-AA013910) that exhibit activity elating to  
CC cytokine, cell proliferation or cell differentiation or which may induce  
CC production of other cytokines in other cell populations. The  
CC polynucleotides and polypeptides are useful in gene therapy, vaccines or  
CC peptide therapy. The polypeptides have various cytokine-like activities,  
CC e.g. stem cell growth factor activity, haematopoiesis regulating  
CC activity, tissue growth factor activity, immunomodulatory activity and  
CC activin/inhibin activity and may be useful in the diagnosis and/or  
CC treatment of cancer, leukaemia, nervous system disorders, arthritis and  
CC inflammation.  
CC Note: The sequence data for this patent did not form part of the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at [ftp.wipo.int/pub/published\\_pcr-sequences](http://wipo.int/pub/published_pcr-sequences).  
XX  
SQ Sequence 34 AA:  
  
Alignment Scores:  
Pred. No.: 1.53 Length: 34  
Score: 10.00 Matches: 10  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatch: 0  
Query Match: 2.10% Indels: 0  
DB: 22 Gaps: 0  
  
US-10-020-441-1 (1-1477) x AA009392 (1-34)  
QY 232 GCGTGCCTGCTTGTGTTGTTGTTGTTATGC 261  
DB 9 AAcysAlaCysValCysValCysValCys 18  
  
RESULT 5  
ID AAM21484  
AAW21484 standard; peptide; 11 AA.  
XX  
AC AAM21484;  
XX  
DT 30-JUL-1997 (first entry)  
XX  
DE Schistosoma elastase precursor derived signal oligopeptide #4.

XX  
KW Hydrophilic; signal oligopeptide; hydrophilicity maxima; vaccine; SIV;  
KW competitive inhibitor; feedback regulator; synthesis; gastrin precursor;  
KW charge; polarity; farnesyl synthetase; plasminogen activator inhibitor 1;  
KW hydroxymethylglutaryl coenzyme A reductase; glucagon precursor; rhesus;  
KW gonadoliberin precursor; plasminogen activator inhibitor 2; prorenin;  
KW Alzheimer amyloid A4; corticotropin releasing factor binding protein;  
KW herpes virus 2 glycoprotein B; HSV2; collagenase; apolipoprotein A;  
KW Treponema pallidum membrane protein; WPA; islet amyloid polypeptide;  
KW fibroblast MMP1; schistosoma elastase precursor; schistosomiasis;  
KW hepatitis delta antigen; rev protein; HIV; VILV; angiotensinogen.  
XX  
OS Schistosoma sp.  
XX  
PN W09519568-A1.  
XX  
PD 20-JUL-1995.  
XX  
PF 12-JAN-1995; 95WO-US00575.  
XX  
PR 14-JAN-1994; 94US-0182248.  
XX  
PA (RATH/) RATH M.  
XX  
PI Rath M;  
XX  
DR WPI: 1995-263953/34.  
XX  
PT Identifying signal oligopeptide(s) in protein sequence(s) - shown as  
PT regions of max. hydrophilicity, used in modulating communication  
PT between protein(s)  
XX  
PS Claim 5; Page 70; 88pp; English.  
XX  
CC The sequences given in AAM21201-560 represent hydrophilic signal oligo-  
CC peptides. These signal oligopeptides are localised on the surface  
CC of the protein and are represented by the hydrophilicity maxima of  
CC the protein. These peptides are enriched in charged amino acids  
CC arranged with neutral spacer amino acids. The specific signal  
CC character of these oligopeptides is determined by a characteristic  
CC combination of conformation and charge within the signal sequence.  
CC These oligopeptides may be used as vaccines in the treatment of  
CC human disease, as competitive inhibitors to prevent or reduce the  
CC metabolic action or interaction of a selected protein by blocking  
CC its specific signal sequences, or as therapeutic agents to function  
CC as feedback regulators to reduce synthesis rate of a selected protein.  
CC These peptides may be modified by omitting one or more amino acids at  
CC the N- and/or C-terminal, by substituting one or more amino acids  
CC without consideration of charge and polarity, by substituting one or  
CC more amino acids with amino acid residues with similar charge and/or  
CC polarity, by omitting one or more amino acids or a combination of these.  
XX  
SQ Sequence 11 AA:  
  
Alignment Scores:  
Pred. No.: 15.3 Length: 11  
Score: 9.00 Matches: 9  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatch: 0  
Query Match: 1.89% Indels: 0  
DB: 16 Gaps: 0  
  
US-10-020-441-1 (1-1477) x AAM21484 (1-11)  
QY 574 AGGATGATACGACCGTGATCGTCA 600  
DB 1 ArgAspAspAsnAspArgAspProSer 9  
  
RESULT 6  
ID AA002434  
AA002434 standard; Protein; 40 AA.  
XX

CC	AAO02434.
XX	
DT	06-NOV-2001 (first entry)
XX	
DE	Human polypeptide SEQ ID NO 16326.
XX	
KW	Human: cytokine; cell proliferation; cell differentiation; gene therapy;
KM	vaccine; peptide therapy; stem cell growth factor; haematopoiesis;
KW	tissue growth factor; immunomodulatory; cancer; leukaemia;
KM	nervous system disorders; arthritis; inflammation.
XX	
OS	Homo sapiens.
PN	MO200164835-A2.
XX	
PD	07-SEP-2001.
XX	
PF	26-FEB-2001; 2001WO-US04927.
XX	
PR	28-FEB-2000; 2000US-0515126.
PR	18-MAY-2000; 2000US-0577409.
XX	
PA	(HYSE-) HYSEQ INC.
XX	
PI	Tang YT, Liu C, Drmanac RT;
DR	WPI; 2001-514838/56.
DR	N-Psdb; AA182365.
PT	Isolated nucleic acids and polypeptides, useful for preventing
PT	diagnosing and treating e.g. leukaemia, inflammation and immune
PT	disorders -
XX	
PS	Claim 20; SEQ ID NO 16326; 1399pp + Sequence Listing; English.
XX	
CC	The invention relates to human polynucleotides (AA179941-AA193841) and
CC	the encoded proteins (AA000010-AA013910) that exhibit activity elating to
CC	cytokine, cell proliferation or cell differentiation or which may induce
CC	production of other cytokines in other cell populations. The
CC	polynucleotides and polypeptides are useful in gene therapy, vaccines or
CC	peptide therapy. The polypeptides have various cytokine-like activities,
CC	e.g. stem cell growth factor activity, haematopoiesis regulating
CC	activity, tissue growth factor activity, immunomodulatory activity and
CC	activin/inhibin activity and may be useful in the diagnosis and/or
CC	treatment of cancer, leukaemia, nervous system disorders, arthritis and
CC	Inflammation.
CC	Note: The sequence data for this patent did not form part of the printed
CC	specification, but was obtained in electronic format directly from WIPO
CC	at ftp.wipo.int/pub/published_pct_sequences.
XX	
SQ	Sequence 40 AA;
XX	
Alignment Scores:	
Pred. No.:	13 6
Score:	9.00 Matches: 9 Length: 40
Percent Similarity:	100.00% Conservative: 0
Best Local Similarity:	100.00% Mismatches: 0
Query Match:	1.89% Indels: 0
DB:	22 Gaps: 0
US-10-020-441-1 (1-1477) x AAO02434 (1-40)	
OY	235 TGTGCGTGCCTTTGTGTTGTATGC 261
Db	24 CysAlaCysValcysValcysValcys 32
RESULT 7	
ID	AAI64770 standard; Protein; 44 AA.
AC	AAI64770;
XX	
DT	01-FEB-2000 (first entry)

XX	Human 5' EST related polypeptide SEQ ID NO:931.
DE	
KW	Human; 5' EST; expressed sequence tag; secreted protein; diagnosis;
KM	gene therapy; chromosome mapping; upstream regulatory sequence;
KW	forensic; location; development; protein synthesis; stability;
KM	regulation; identification.
XX	
OS	Homo sapiens.
PN	MO9953051-A2.
PD	21-OCT-1999.
XX	
PF	09-APR-1999; 99WO-IB00712.
XX	
PR	09-APR-1998; 98US-0057719.
PA	28-APR-1998; 98US-0069047.
XX	
(GEST ) GENSET.	
Dumas Milne Edwards J, Duclert A, Giordano J;	
WP1: 2000-038446/03.	
N-PDB: AA242384.	
Novel secreted protein 5' expressed sequence tag sequences used in diagnostic, forensic, gene therapy, and chromosome mapping procedures	
Claim 3; Page 637; 837pp; English.	
AA242265 to AA243075 represent novel 5' expressed sequence tag (EST) sequences, corresponding to human secreted proteins. AA242265 to AA243075 represent the EST-related proteins corresponding to AA242265 to AA243075. The 5' ESTs can be used for producing secreted human gene products. They can be used to identify and isolate 5' untranslated regions (UTRs), and upstream regulatory regions which control the location, development stage, rate, and quantity of protein synthesis, as well as stability of mRNA. The ESTs are also useful as probes for chromosome mapping, and to obtain full length cDNA clones. The ESTs can also be used in forensic procedures to identify individuals, or in diagnostic procedures to identify individuals having genetic diseases resulting from abnormal gene expression. The products may also be used in gene therapy protocols. The nucleic acids encoding signal peptides can be used for directing extracellular secretion of a polypeptide or the insertion of a polypeptide into a membrane, or importing a polypeptide into a cell. The proteins encoded by the EST sequences may be useful in treating a variety of human conditions. Secreted proteins have therapeutic value, and the identification of new secreted proteins is valuable. AA242249 to AA242264 and AA244644 to AA244650 represent sequences used in the exemplification of the present invention.	
Sequence	44 AA;
Alignment Scores:	
Pred. No.:	13.4
Score:	9.00
Percent Similarity:	100.00%
Best local Similarity:	100.00%
Query Match:	1.89%
Gaps:	0
Length:	44
Matches:	9
Conservative:	0
Mismatches:	0
Indels:	0
US-10-020-441-1 (1-1477) x AA244770 (1-44)	
OY	694 GTCGTGATCTGTTGTCTGCTGC 720       Db 14 VALCYSVALSERVALSERVAL 22
RESULT 8	
ID	AA007568 standard; Protein; 55 AA.
CC	AA007568;

XX 06-NOV-2001 (first entry)  
 XX  
 XX Human polypeptide SEQ ID NO 21460.  
 DE  
 XX  
 XX Human: cytokine; cell proliferation; cell differentiation; gene therapy;  
 KW vaccine; peptide therapy; stem cell growth factor; hematopoiesis;  
 KW tissue growth factor; immunomodulatory; cancer; leukemia;  
 KW nervous system disorders; arthritis; inflammation.  
 XX  
 OS Homo sapiens.  
 XX  
 XX WO200164835-A2.  
 XX  
 XX 07-SEP-2001.  
 XX  
 XX 26-FEB-2001; 2001WO-US04927.  
 XX  
 XX 28-FEB-2000; 2000US-0515126.  
 XX 18-MAY-2000; 2000US-0577409.  
 XX  
 XX (HYSE-) HYSEQ INC.  
 XX  
 XX Tang YT, Liu C, Drmanac RT;  
 XX  
 XX WPI: 2001-514838/56.  
 XX N-PSDB: AAI87499.  
 XX  
 XX Isolated nucleic acids and polypeptides, useful for preventing  
 PT diagnosing and treating e.g. leukemia, inflammation and immune  
 PT disorders -  
 XX  
 PS Claim 20: SEQ ID NO 21460; 1399pp + Sequence Listing; English.  
 XX  
 CC The invention relates to human polynucleotides (AAI79941-AAI93941) and  
 CC the encoded proteins (AAO00010-AAO13910) that exhibit activity elating to  
 CC cytokine, cell proliferation or cell differentiation or which may induce  
 CC production of other cytokines in other cell populations. The  
 CC polynucleotides and polypeptides are useful in gene therapy, vaccines or  
 CC peptide therapy. The polypeptides have various cytokine-like activities,  
 CC e.g. stem cell growth factor activity, hematopoiesis regulating  
 CC activity, tissue growth factor activity, immunomodulatory activity and  
 CC activin/inhibin activity and may be useful in the diagnosis and/or  
 CC treatment of cancer, leukemia, nervous system disorders, arthritis and  
 CC inflammation.  
 CC Note: The sequence data for this patent did not form part of the printed  
 CC specification, but was obtained in electronic format directly from WIPO  
 CC at ftp.wipo.int/pub/published\_pcl\_sequences.  
 CC  
 XX Sequence 55 AA;  
 SQ  
 Alignment Scores:  
 Pred. No.: 13.2 Length: 55  
 Score: 9.00 Matches: 9  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 1.89% Indels: 0  
 DB: 22 Gaps: 0  
 US-10-020-441-1 (1-1477) x AAO07568 (1-55)  
 OY 235 TGTGCGTGGCTTGTGTATGC 261  
 DB 24 CysAlaCysAlaCysAlaCys 32  
 RESULT 9  
 ID ABP00091 standard; Protein: 56 AA.  
 XX  
 AC ABP00091;  
 XX  
 XX 24-JUN-2002 (first entry)  
 DT

DE Human ORFX protein sequence SEQ ID NO:164.  
 XX  
 XX Human: open reading frame; ORFX; gene therapy; cancer; cirrhosis;  
 KW hyperproliferative disorder; psoriasis; benign tumour; haemorrhage;  
 KW degenerative disorder; osteoarthritis; neurodegenerative disorder;  
 KW cardiovascular disease; diabetes mellitus; systemic lupus erythematosus;  
 KW hypertension; hypothyroidism; cholesterol ester storage disease;  
 KW immune deficiency; immune disorder; infectious disease;  
 KW autoimmune disorder; rheumatoid arthritis; autoimmune thyroiditis;  
 KW myasthenia gravis.  
 XX  
 OS Homo sapiens.  
 XX  
 XX WO200192523-A2.  
 XX  
 XX 06-DEC-2001.  
 XX  
 XX 29-MAY-2001; 2001WO-US10836.  
 XX  
 XX 30-MAY-2000; 2000US-206132P.  
 XX 29-AUG-2000; 2000US-228716P.  
 XX  
 XX (CURA-) CURAGEN CORP.  
 XX  
 XX Shimkets RA, Leach MD;  
 XX  
 XX WPI: 2002-106308/14.  
 XX N-PSDB: ABN15843.  
 XX  
 XX Novel human polypeptides and polynucleotides useful for diagnosing,  
 PT preventing and treating cardiovascular disease, neurodegenerative,  
 PT hyperproliferative disorders and autoimmune disorders -  
 XX  
 PS Disclosure: SEQ ID 164; 1037pp; English.  
 XX  
 CC The present invention describes substantially purified human proteins  
 CC (referred to as open reading frame, ORFX, where X is 1-11491 (see Table 1  
 CC in the specification). ABN15762 to ABN27252 encode the human ORFX  
 CC proteins given in ABP00010 to ABP11500. ORFX proteins are useful for  
 CC treating or preventing a pathology associated with an ORFX-associated  
 CC disorder in humans, and in the manufacture of a medicament for treating a  
 CC syndrome associated with ORFX-associated disorder. ORFX polynucleotide  
 CC sequences can be used in gene therapy. ORFX sequences can be used in the  
 CC treatment of cancer, hyperproliferative disorders, cirrhosis of liver,  
 CC psoriasis, benign tumours, keloid, degenerative disorders, haemorrhage,  
 CC osteoarthritis, neurodegenerative disorders, disorders related to organ  
 CC transplantation, cardiovascular diseases, diabetes mellitus, systemic  
 CC lupus erythematosus, hypertension, hypothyroidism, cholesterol ester  
 CC storage disease, various immune deficiencies and disorders, infectious  
 CC diseases, autoimmune disorders such as multiple sclerosis, rheumatoid  
 CC arthritis, autoimmune thyroiditis, myasthenia gravis, graft-versus-host  
 CC disease and autoimmune inflammatory eye disease. ORFX proteins are also  
 CC useful for treating burns, incisions, ulcers, for treating osteoporosis,  
 CC bone degenerative disorders, or periodontal disease, and for gut  
 CC protection or regeneration and treatment of lung or liver fibrosis,  
 CC reperfusion injury in various tissues and conditions resulting from  
 CC systemic cytokine damage.  
 CC N.B. The sequence data for this patent did not form part of the printed  
 CC specification, but was obtained in electronic format directly from WIPO  
 CC at ftp.wipo.int/pub/published\_pcl\_sequences.  
 CC  
 XX Sequence 56 AA;  
 SQ  
 Alignment Scores:  
 Pred. No.: 13.1 Length: 56  
 Score: 9.00 Matches: 9  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 1.89% Indels: 0  
 DB: 23 Gaps: 0  
 US-10-020-441-1 (1-1477) x ABP00091 (1-56)  
 ;

QY 691 TGTGCTGTATCTGTTGTGTCT 717  
 ||||||||||||||||||  
 Db 38 CysValCysValSerValCysValSer 46

RESULT 10  
 ABG24997  
 ID ABG24997 standard; Protein: 922 AA.  
 XX  
 AC ABG24997;  
 XX  
 DT 18-FEB-2002 (first entry)  
 XX  
 DE Novel human diagnostic protein #24988.  
 XX  
 KW Human: chromosome mapping; gene therapy; forensic;  
 KW food supplement; medical imaging; diagnostic; genetic disorder.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO200175067-A2.  
 XX  
 PD 11-OCT-2001.  
 XX  
 PF 30-MAR-2001; 2001WO-US08631.  
 XX  
 PR 31-MAR-2000; 2000US-0540217.  
 PR 23-AUG-2000; 2000US-0649167.  
 XX  
 PA (HYSE-) HYSEQ INC.  
 PI Drmanac RT, Liu C, Tang YT;  
 PI WPI: 2001-639362/73.  
 DR N-PSDB: AAS89184.  
 DR  
 XX  
 PT New isolated polynucleotide and encoded polypeptides, useful in  
 PT diagnostics, forensics, gene mapping, identification of mutations  
 PT responsible for genetic disorders or other traits and to assess  
 PT biodiversity -  
 PS  
 PS Claim 20: SEQ ID NO 55356; 103pp: English.  
 XX  
 CC The invention relates to isolated polynucleotide (I) and  
 CC polypeptide (II) sequences. (I) is useful as hybridisation probes,  
 CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome  
 CC and gene mapping, and in recombinant production of (II). The  
 CC polynucleotides are also used in diagnostics as expressed sequence tags  
 CC for identifying expressed genes. (I) is useful in gene therapy techniques  
 CC to restore normal activity of (II) or to treat disease states involving  
 CC (II). (II) is useful for generating antibodies against it, detecting or  
 CC quantitating a polypeptide in tissue, as molecular weight markers and as  
 CC a food supplement. (II) and its binding partners are useful in medical  
 CC imaging of sites expressing (II). (I) and (II) are useful for treating  
 CC disorders involving aberrant protein expression or biological activity.  
 CC The polypeptide and polynucleotide sequences have applications in  
 CC diagnostics, forensics, gene mapping, identification of mutations  
 CC responsible for genetic disorders or other traits to assess biodiversity  
 CC and to produce other types of data and products dependent on DNA and  
 CC amino acid sequences. ABG00010-ABG30377 represent novel human  
 CC diagnostic amino acid sequences of the invention.  
 CC Note: The sequence data for this patent did not appear in the printed  
 CC specification, but was obtained in electronic format directly from WIPO  
 CC at ftp.wipo.int/pub/published\_pct\_sequences.  
 CC  
 XX  
 SO Sequence 922 AA:

Alignment Scores:  
 Pred. No.: 10.1 Length: 922  
 Score: 9.00 Matches: 9  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 1.89% Indels: 0  
 DB: 22 Gaps: 0

US-10-020-441-1 (1-1477) x ABG24997 (1-922)

QY 1430 TATGCCACTGTGCTAGAAATGTTGGAT 1456  
 ||||||||||||||||||  
 Db 851 TyraIaSerValAlaArgMetLeuasp 859

RESULT 11  
 AA005400  
 ID AA005400 standard; Protein: 28 AA.  
 XX  
 AC AA005400;  
 XX  
 DT 06-NOV-2001 (first entry)  
 XX  
 DE Human polypeptide SEQ ID NO 19292.  
 XX  
 KW Human: cytokine; cell proliferation; cell differentiation; gene therapy;  
 KW vaccine; peptide therapy; stem cell growth factor; haematopoiesis;  
 KW tissue growth factor; immunomodulatory; cancer; leukaemia;  
 KW nervous system disorders; arthritis; inflammation.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO200164835-A2.  
 XX  
 PD 07-SEP-2001.  
 XX  
 PF 26-FEB-2001; 2001WO-US04927.  
 XX  
 PR 28-FEB-2000; 2000US-0515126.  
 PR 18-MAY-2000; 2000US-0577409.  
 XX  
 PA (HYSE-) HYSEQ INC.  
 PI Tang YT, Liu C, Drmanac RT;  
 PI WPI: 2001-514838/56.  
 DR N-PSDB: AAI85331.  
 DR  
 XX  
 PT Isolated nucleic acids and polypeptides, useful for preventing  
 PT diagnosing and treating e.g. leukaemia, inflammation and immune  
 PT disorders -  
 PS  
 PS Claim 20: SEQ ID NO 19292; 1399pp + Sequence Listing: English.  
 XX  
 CC The invention relates to human polynucleotides (AAI79941-AAI93841) and  
 CC the encoded proteins (AA000010-AA013910) that exhibit activity relating to  
 CC cytokine, cell proliferation or cell differentiation or which may induce  
 CC production of other cytokines in other cell populations. The  
 CC polynucleotides and polypeptides are useful in gene therapy, vaccines or  
 CC peptide therapy. The polypeptides have various cytokine-like activities,  
 CC e.g. stem cell growth factor activity, haematopoiesis regulating  
 CC activity, tissue growth factor activity, immunomodulatory activity and  
 CC activin/inhibin activity and may be useful in the diagnosis and/or  
 CC treatment of cancer, leukaemia, nervous system disorders, arthritis and  
 CC inflammation.  
 CC Note: The sequence data for this patent did not form part of the printed  
 CC specification, but was obtained in electronic format directly from WIPO  
 CC at ftp.wipo.int/pub/published\_pct\_sequences.  
 CC  
 XX  
 SO Sequence 28 AA:

Alignment Scores:  
 Pred. No.: 126 Length: 28  
 Score: 8.00 Matches: 8  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 1.68% Indels: 0  
 DB: 22 Gaps: 0

US-10-020-441-1 (1-1477) x AA005400 (1-28)

OY 241 TCGCTTGTGTGTGATGCAGT 264  
Db 7 CysValCysValCysValCysSer 14  
RESULT 12  
ID AAO04534 standard; Protein: 29 AA.  
XX AAO04534;  
AC AAO04534;  
XX 06-NOV-2001 (first entry)  
DT  
XX Human polypeptide SEQ ID NO 18426.  
DE  
XX Human: cytokine; cell proliferation; cell differentiation; gene therapy;  
KW vaccine; peptide therapy; stem cell growth factor; haematopoiesis;  
KM tissue growth factor; immunomodulatory; cancer; leukaemia;  
XX nervous system disorders; arthritis; inflammation.  
OS Homo sapiens.  
XX WO200164835-A2.  
PN  
XX 07-SEP-2001.  
PD  
PF 26-FEB-2001; 2001MO-US04927.  
XX  
XX 28-FEB-2000; 2000US-0515126.  
PR 18-MAY-2000; 2000US-0577409.  
XX  
XX (HYSE-) HYSEQ INC.  
PA  
XX  
XX Tang YT, Liu C, Drmanac RT;  
PI WPI; 2001-514838/56.  
XX  
XX N-PSDB; AA184465.  
DR  
XX Isolated nucleic acids and polypeptides, useful for preventing  
PT diagnosing and treating e.g. leukaemia, inflammation and immune  
PT disorders -  
XX  
XX Claim 20; SEQ ID NO 18426; 1399pp + Sequence Listing; English.  
XX  
XX The invention relates to human polynucleotides (AA179941-AA193841) and  
CC the encoded proteins (AA000010-AA013910) that exhibit activity elating to  
CC cytokine, cell proliferation or cell differentiation or which may induce  
CC production of other cytokines in other cell populations. The  
CC polynucleotides and polypeptides are useful in gene therapy, vaccines or  
CC peptide therapy. The polypeptides have various cytokine-like activities,  
CC e.g. stem cell growth factor activity, haematopoiesis regulating  
CC activity, tissue growth factor activity, immunomodulatory activity and  
CC activin/inhibin activity and may be useful in the diagnosis and/or  
CC treatment of cancer, leukaemia, nervous system disorders, arthritis and  
CC inflammation.  
CC Note: The sequence data for this patent did not form part of the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pct\_sequences.  
XX  
XX SQ Sequence 29 AA;  
XX  
XX Alignment Scores:  
Pred. No.: 126 Length: 29  
Score: 8.00 Matches: 8  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 1.68% Indels: 0  
DB: 22 Gaps: 0  
US-10-020-441-1 (1-1477) x AAO04534 (1-29)  
OY 238 GCGTGCCTTGTGTGTGATGC 261  
Db 11 AlacysValCysValCysValCys 18

RESULT 13  
ID AAG58999 standard; Protein: 36 AA.  
XX AAG58999;  
AC AAG58999;  
XX 18-OCT-2000 (first entry)  
DT  
XX Arabidopsis thaliana protein fragment SEQ ID NO: 76268.  
DE  
XX Protein identification; signal transduction pathway; metabolic pathway;  
KW hybridisation assay; genetic mapping; gene expression control; promoter;  
KM termination sequence.  
XX  
XX Arabidopsis thaliana.  
OS  
XX EPI033405-A2.  
PN  
XX 06-SEP-2000.  
PD  
XX  
XX 25-FEB-2000; 2000EP-0301439.  
PF  
XX  
XX 25-FEB-1999; 99US-0121825.  
PR 05-MAR-1999; 99US-0123180.  
PR 03-MAR-1999; 99US-0123348.  
PR 23-MAR-1999; 99US-0125788.  
PR 25-MAR-1999; 99US-0126264.  
PR 29-MAR-1999; 99US-0126785.  
PR 01-APR-1999; 99US-0127462.  
PR 06-APR-1999; 99US-0128234.  
PR 08-APR-1999; 99US-0128714.  
PR 16-APR-1999; 99US-0129845.  
PR 19-APR-1999; 99US-0130077.  
PR 23-APR-1999; 99US-0130449.  
PR 23-APR-1999; 99US-0130510.  
PR 23-APR-1999; 99US-0130891.  
PR 28-APR-1999; 99US-0131449.  
PR 30-APR-1999; 99US-0132048.  
PR 30-APR-1999; 99US-0132407.  
PR 04-MAY-1999; 99US-0132484.  
PR 05-MAY-1999; 99US-0132485.  
PR 06-MAY-1999; 99US-0132486.  
PR 06-MAY-1999; 99US-0132487.  
PR 07-MAY-1999; 99US-0132863.  
PR 11-MAY-1999; 99US-0134256.  
PR 14-MAY-1999; 99US-0134218.  
PR 14-MAY-1999; 99US-0134219.  
PR 14-MAY-1999; 99US-0134221.  
PR 14-MAY-1999; 99US-0134370.  
PR 18-MAY-1999; 99US-0134768.  
PR 19-MAY-1999; 99US-0134941.  
PR 20-MAY-1999; 99US-0135124.  
PR 21-MAY-1999; 99US-0135353.  
PR 24-MAY-1999; 99US-0135629.  
PR 25-MAY-1999; 99US-0136021.  
PR 27-MAY-1999; 99US-0136392.  
PR 28-MAY-1999; 99US-0136782.  
PR 01-JUN-1999; 99US-0137222.  
PR 03-JUN-1999; 99US-0137528.  
PR 04-JUN-1999; 99US-0137502.  
PR 07-JUN-1999; 99US-0137724.  
PR 08-JUN-1999; 99US-0138094.  
PR 10-JUN-1999; 99US-0138540.  
PR 10-JUN-1999; 99US-0138847.  
PR 14-JUN-1999; 99US-0139119.  
PR 16-JUN-1999; 99US-0139452.  
PR 16-JUN-1999; 99US-0139453.  
PR 17-JUN-1999; 99US-0139492.  
PR 18-JUN-1999; 99US-0139454.  
PR 18-JUN-1999; 99US-0139455.  
PR 18-JUN-1999; 99US-0139456.  
PR 18-JUN-1999; 99US-0139457.

PR 18-JUN-1999; 99US-0139458.  
PR 18-JUN-1999; 99US-0139459.  
PR 18-JUN-1999; 99US-0139460.  
PR 18-JUN-1999; 99US-0139461.  
PR 18-JUN-1999; 99US-0139462.  
PR 18-JUN-1999; 99US-0139463.  
PR 18-JUN-1999; 99US-0139750.  
PR 18-JUN-1999; 99US-0139751.  
PR 21-JUN-1999; 99US-0139817.  
PR 22-JUN-1999; 99US-0139889.  
PR 23-JUN-1999; 99US-0140353.  
PR 23-JUN-1999; 99US-0140354.  
PR 24-JUN-1999; 99US-0140695.  
PR 28-JUN-1999; 99US-0140823.  
PR 29-JUN-1999; 99US-0140991.  
PR 30-JUN-1999; 99US-0141287.  
PR 01-JUL-1999; 99US-0141842.  
PR 01-JUL-1999; 99US-0142154.  
PR 02-JUL-1999; 99US-0142055.  
PR 06-JUL-1999; 99US-0142390.  
PR 08-JUL-1999; 99US-0142803.  
PR 09-JUL-1999; 99US-0142920.  
PR 12-JUL-1999; 99US-0142977.  
PR 13-JUL-1999; 99US-0143542.  
PR 14-JUL-1999; 99US-0143624.  
PR 15-JUL-1999; 99US-0144005.  
PR 16-JUL-1999; 99US-0144085.  
PR 16-JUL-1999; 99US-0144086.  
PR 19-JUL-1999; 99US-0144325.  
PR 19-JUL-1999; 99US-0144331.  
PR 19-JUL-1999; 99US-0144332.  
PR 19-JUL-1999; 99US-0144333.  
PR 19-JUL-1999; 99US-0144334.  
PR 19-JUL-1999; 99US-0144335.  
PR 20-JUL-1999; 99US-0144352.  
PR 20-JUL-1999; 99US-0144632.  
PR 20-JUL-1999; 99US-0144684.  
PR 21-JUL-1999; 99US-0144814.  
PR 21-JUL-1999; 99US-0145086.  
PR 21-JUL-1999; 99US-0145088.  
PR 22-JUL-1999; 99US-0145085.  
PR 22-JUL-1999; 99US-0145087.  
PR 22-JUL-1999; 99US-0145089.  
PR 22-JUL-1999; 99US-0145192.  
PR 23-JUL-1999; 99US-0145145.  
PR 23-JUL-1999; 99US-0145218.  
PR 23-JUL-1999; 99US-0145224.  
PR 26-JUL-1999; 99US-0145276.  
PR 27-JUL-1999; 99US-0145913.  
PR 27-JUL-1999; 99US-0145919.  
PR 28-JUL-1999; 99US-0145951.  
PR 02-AUG-1999; 99US-0146386.  
PR 02-AUG-1999; 99US-0146388.  
PR 02-AUG-1999; 99US-0146389.  
PR 03-AUG-1999; 99US-0147038.  
PR 04-AUG-1999; 99US-0147204.  
PR 04-AUG-1999; 99US-0147302.  
PR 05-AUG-1999; 99US-0147192.  
PR 05-AUG-1999; 99US-0147260.  
PR 06-AUG-1999; 99US-0147303.  
PR 06-AUG-1999; 99US-0147416.  
PR 09-AUG-1999; 99US-0147493.  
PR 09-AUG-1999; 99US-0147935.  
PR 10-AUG-1999; 99US-0148171.  
PR 11-AUG-1999; 99US-0148319.  
PR 12-AUG-1999; 99US-0148341.  
PR 13-AUG-1999; 99US-0148565.  
PR 13-AUG-1999; 99US-0148684.  
PR 16-AUG-1999; 99US-0149358.  
PR 17-AUG-1999; 99US-0149175.  
PR 18-AUG-1999; 99US-0149426.  
PR 20-AUG-1999; 99US-0149722.

PR 20-AUG-1999; 99US-0149723.  
PR 20-AUG-1999; 99US-0149929.  
PR 23-AUG-1999; 99US-0149902.  
PR 23-AUG-1999; 99US-0149930.  
PR 25-AUG-1999; 99US-0150566.  
PR 26-AUG-1999; 99US-0150884.  
PR 27-AUG-1999; 99US-0151065.  
PR 27-AUG-1999; 99US-0151066.  
PR 27-AUG-1999; 99US-0151080.  
PR 30-AUG-1999; 99US-0151303.  
PR 31-AUG-1999; 99US-0151303.  
PR 01-SEP-1999; 99US-0151930.  
PR 07-SEP-1999; 99US-0152363.  
PR 10-SEP-1999; 99US-0153070.  
PR 13-SEP-1999; 99US-0153758.  
PR 15-SEP-1999; 99US-0154018.  
PR 16-SEP-1999; 99US-0154039.  
PR 20-SEP-1999; 99US-0154779.  
PR 22-SEP-1999; 99US-0155139.  
PR 23-SEP-1999; 99US-0155486.  
PR 24-SEP-1999; 99US-0155659.  
PR 28-SEP-1999; 99US-0156458.  
PR 29-SEP-1999; 99US-0156596.  
PR 04-OCT-1999; 99US-0157117.  
PR 05-OCT-1999; 99US-0157753.  
PR 06-OCT-1999; 99US-0157865.  
PR 07-OCT-1999; 99US-0158029.  
PR 08-OCT-1999; 99US-0158232.  
PR 12-OCT-1999; 99US-0158369.  
PR 13-OCT-1999; 99US-0159293.  
PR 13-OCT-1999; 99US-0159294.  
PR 13-OCT-1999; 99US-0159295.  
PR 14-OCT-1999; 99US-0159329.  
PR 14-OCT-1999; 99US-0159330.  
PR 14-OCT-1999; 99US-0159331.  
PR 14-OCT-1999; 99US-0159637.  
PR 14-OCT-1999; 99US-0159638.  
PR 18-OCT-1999; 99US-0159584.  
PR 21-OCT-1999; 99US-0160741.  
PR 21-OCT-1999; 99US-0160767.  
PR 21-OCT-1999; 99US-0160768.  
PR 21-OCT-1999; 99US-0160770.  
PR 21-OCT-1999; 99US-0160814.  
PR 21-OCT-1999; 99US-0160815.  
PR 22-OCT-1999; 99US-0160980.  
PR 22-OCT-1999; 99US-0160981.  
PR 22-OCT-1999; 99US-0160989.  
PR 25-OCT-1999; 99US-0161404.  
PR 25-OCT-1999; 99US-0161405.  
PR 25-OCT-1999; 99US-0161406.  
PR 26-OCT-1999; 99US-0161359.  
PR 26-OCT-1999; 99US-0161360.  
PR 26-OCT-1999; 99US-0161361.  
PR 28-OCT-1999; 99US-0161920.  
PR 28-OCT-1999; 99US-0161992.  
PR 28-OCT-1999; 99US-0161993.  
PR 29-OCT-1999; 99US-0162142.

## Alignment Scores:

Pred. No.:	123	Length:	36
Score:	8.00	Matches:	8
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	1.66%	Indels:	0
DB:	21	Gaps:	0

US-10-020-441-1 (1-1477) x AAG58999 (1-36)

QY 11A3 TTTTGTTCCTTCCTGTGTTT 1166

Db 10 PhcysphevalProlcucysphe 17

RESULT 14

AAB44943  
ID AAB44943 standard; Protein: 47 AA.  
XX  
XX AAB44943;  
XX  
DT 12-FEB-2001 (first entry)  
XX  
DE Human secreted protein encoded by gene 27.  
XX  
XX Secreted protein; human; immunosuppressive; antiarthritic; antirheumatic;  
KM antiproliferative; cytostatic; cardiant; vasotropic; cerebroprotective;  
KM neurotropic; neuroprotective; antibacterial; vituicide; fungicide;  
KM ophthalmological; gene therapy; treatment; autoimmune disease; infection;  
KM cardiovascular disorder; nervous system disorder; ocular disorder;  
KM wound healing; epithelial cell proliferation; skin aging; mental state;  
transplantation; metabolism modulation.  
XX  
OS Homo sapiens.  
XX  
PN WO200055200-A1.  
XX  
PD 21-SEP-2000.  
XX  
PF 09-MAR-2000; 2000WO-US06042.  
XX  
PR 12-MAR-1999; 99US-0124143.  
PR 03-DEC-1999; 99US-0168663.  
XX  
PA (HUMA-) HUMAN GENOME SCI INC.  
XX  
PI Rosen CA, Ruben SM, Komatsoulis G;  
XX  
DR WPI; 2000-656008/63.  
DR N-PSDB: AAC/9924.  
XX  
XX Isolated human secretory proteins, nucleic acids encoding them and  
PT antibodies directed against them, useful for diagnosing and treating  
PT disorders related to the proteins such as cancer, Alzheimer's disease  
PT and Parkinsons -  
XX  
XX  
PS Claim 11; Page 390; 453pp; English.  
XX  
XX This invention describes a novel isolated polypeptide (I) and its  
CC encoding nucleic acid molecule (II) which have immunosuppressive,  
CC antiarthritic, antirheumatic, antiproliferative, cytostatic, cardiant,  
CC vasotropic, cerebroprotective, neurotropic, neuroprotective, antibacterial,  
CC virucide, fungicide and ophthalmological activity and which can be used  
CC for gene therapy. (I) and (II) are used to prevent, treat or ameliorate  
CC a medical condition in e.g. humans, mice, rabbits, goats, horses, cats,  
CC dogs, chickens or sheep. (I) and (II) are also used in diagnosing a  
CC pathological condition or susceptibility to a pathological condition. The  
CC antibodies to (II) can also be used in alleviating symptoms associated  
CC with the disorders and in diagnostic immunoassays e.g. radioimmunoassays  
CC or enzyme linked immunosorbent assays (ELISA). Disorders which are  
CC diagnosed or treated include autoimmune diseases e.g. rheumatoid  
CC arthritis, hyperproliferative disorders e.g. neoplasms of the breast or  
CC liver, cardiovascular disorders e.g. cardiac arrest, cerebrovascular  
CC disorders e.g. cerebral ischemia, angiogenesis, nervous system disorders  
CC e.g. Alzheimer's disease, infections caused by bacteria, viruses and  
CC fungi and ocular disorders e.g. corneal infection. The polypeptides can  
CC also be used to aid wound healing and epithelial cell proliferation, to  
CC prevent skin aging due to sunburn, to maintain organs before  
CC transplantation, for supporting cell culture of primary tissues, to  
CC regenerate tissues and in chemotaxis. The polypeptides are used to  
CC modulate mammalian metabolism, to change mammal's mental state or  
CC physical state by influencing biorhythms circadian rhythms, depression  
CC tendency for violence tolerance for pain, reproductive capabilities,  
CC hormonal or endocrine levels, appetite, libido, memory, stress or other  
CC cognitive qualities, as a food additive or preservative, such as to  
CC increase or decrease storage capabilities, fat content, lipid, protein,  
CC carbohydrates, vitamins, minerals, cofactors or other nutritional  
XX components.

SQ Sequence 47 AA;  
Alignment Scores:  
Pred. No.: 120 Length: 47  
Score: 8.00 Matches: 8  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 1.68% Indels: 0  
DB: 21 Gaps: 0  
US-10-020-441-1 (1-1477) x AAB44943 (1-47)  
QY 238 GCGTGGCTTGTGTGTGTGTATGC 261  
|||||  
Db 16 AlacysValcysValcysValcys 23  
RESULT 15  
ABP32259  
ID ABP32259 standard; Protein: 53 AA.  
XX  
XX ABP32259;  
XX  
DT 09-JUL-2002 (first entry)  
XX  
DE Human ORF1232 protein, SEQ ID NO:2464.  
XX  
XX Human; ORF; open reading frame; ORF; drug screening; diagnosis;  
KM disease monitoring; cytokine; cell proliferation; cell differentiation;  
KM immune modulation; haematopoiesis regulation; tissue growth;  
KM angiogenesis; activin; inhibin; chemotactic; chemokinetic; haemostatic;  
KM thrombolytic; tumour inhibition; bodily characteristic; fertility;  
KM behaviour; cancer; proliferative disorder; neurological disorder;  
KM cardiovascular disease; immune system disorder; organ transplantation;  
KM tissue growth disorder; tissue regeneration disorder; diabetes mellitus;  
KM hypothyroidism; cholesterol ester storage disease; infection; vulnerability;  
KM vasotropic; antipsoriatic; antidiabetic; cytostatic; neurotropic;  
KM neuroprotective; antiatherosclerotic; anticoagulant; thrombolytic;  
KM cardiant; hypotensive; antithyroid; antiinflammatory; immunomodulator;  
KM dermatological; analgesic; virucide; antibacterial; fungicide.  
XX  
OS Homo sapiens.  
XX  
XX  
PN WO200190366-A2.  
XX  
XX 29-NOV-2001.  
XX  
XX 24-MAY-2001; 2001WO-US17076.  
XX  
XX 24-MAY-2000; 2000US-206690P.  
XX  
XX (CURA-) CURAGEN CORP.  
XX  
PA Leach MD, Shinkets RA;  
XX  
DR WPI; 2002-106200/14.  
DR N-PSDB: ABN/6285.  
XX  
XX Novel human polypeptides and polynucleotides useful for diagnosing,  
PT preventing and treating cardiovascular disease, neurodegenerative,  
PT hyperproliferative disorders and disorders related to organ  
PT transplantation -  
XX  
XX  
PS Claim 10; Page 878; 2508pp; English.  
XX  
XX Sequences ABP31028-ABP35561 represent 4534 novel human proteins  
CC designated ORF (open reading frame) 1-4534, and sequences ABN75054-  
CC ABN79587 represent cDNAs encoding them. The invention also encompasses  
CC polypeptides at least 80% identical to the ORF1-ORF4534 (collectively  
CC referred to as ORFX) proteins, polynucleotides at least 85% identical to  
CC the ORFX nucleic acid sequences, vectors and host cells comprising ORFX  
CC polynucleotides, the recombinant production of ORFX proteins, antibodies  
CC specific for ORFX proteins, methods of detecting ORFX polynucleotides and  
CC polypeptides, methods of screening for modulators of ORFX expression or



CC activity, and methods of screening individuals for a predisposition to an  
 CC ORFX-associated disorder. The ORFX proteins of the invention have a wide  
 CC range of biological activities, such as cytokine, cell proliferation,  
 CC cell differentiation, immune modulation, haematopoiesis regulation,  
 CC tissue growth, angiogenesis, activin or inhibin activity, chemocytic/  
 CC chemokine activity, haemostatic activity, thrombolytic activity,  
 CC receptor/ligand, antiinflammatory activity, tumour inhibition activity,  
 CC and antifibrotic activity, and may also be involved in the determination  
 CC of bodily characteristics, fertility and behaviour. ORFX proteins,  
 CC nucleic acids and antibodies may be used in the treatment of cancers,  
 CC other proliferative disorders such as psoriasis and benign tumours,  
 CC neurological disorders such as epilepsy and Alzheimer's disease,  
 CC cardiovascular diseases, immune system disorders, disorders related to  
 CC organ transplantation, disorders of tissue growth and regeneration,  
 CC diseases such as diabetes mellitus, hypothyroidism, and cholesterol ester  
 CC storage disease, and infectious diseases caused by viral, bacterial,  
 CC fungal and other pathogens. ORFX nucleic acids may also be used as a  
 CC source of primers and probes, in the detection of ORFX genomic sequences  
 CC or transcripts, in the identification and cloning of homologous  
 CC sequences, in genetic diagnosis, and in forensic biology. The ORFX  
 CC nucleic acids may additionally be used to produce transgenic animals  
 CC which may be useful for studying the function and/or activity of ORFX  
 CC protein, and in drug screening. The ORFX proteins may also be used as  
 CC immunogens to generate specific antibodies, which are useful in the  
 CC diagnosis, treatment and monitoring of ORFX-associated diseases.

CC Sequence 53 AA:

Alignment Scores:  
 Pred. No.: 119 Length: 53  
 Score: 8.00 Matches: 8  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 1.68% Indels: 0  
 DB: 23 Gaps: 0

US-10-020-441-1 (1-1477) x ABP32259 (1-53)

Oy 688 CTGTGCTGCTGTAATCTTGT 711  
 ||||||||||||||||||||  
 Db 45 LeuCYsValCYsValSerValCys 52

RESULT 16

AAE01377 standard; Protein: 61 AA.

AC AAE01377;

DT 18-JUL-2001 (first entry)

DE Human gene 3 encoded secreted protein HOHBL42, SEQ ID NO:99.

XX Human; secreted protein; proliferative disorder; cancer; tumour; asthma;  
 KW foetal abnormality; developmental abnormality; haematopoietic disorder;  
 KW immune system disorder; AIDS; autoimmune disease; Rheumatoid arthritis;  
 KW Parkinson's disease; cognitive disorder; schizophrenia; skin disorder;  
 KW psoriasis; sepsis; diabetes; atherosclerosis; cardiovascular disorder;  
 KW inflammation; neurological disorder; Alzheimer's disease; food additive;  
 KW angiogenic disorder; kidney disorder; gastrointestinal disorder; allergy;  
 KW pregnancy-related disorder; endocrine disorder; infection; wound healing;  
 KW cell culture; chemotaxis; vulnery; binding partner identification;  
 KW gene therapy.

XX Homo sapiens.

XX Key Location/Qualifiers

FT Peptide 1..15

FT Protein 16..61

FT Misc-difference 13 Mature\_human\_secreted\_protein

FT /Label= Unknown

FT /note= "Encoded by TRT"

XX MO200134629-A1.

PD 17-MAY-2001.

PF 08-NOV-2000: 2000MO-US030654.

PR 12-NOV-1999: 99US-0164835.

PR 27-JUL-2000: 2000US-0221142.

XX (HUMA-) HUMAN GENOME SCI INC.

PA Ruben SM, Komatsoulis GA, Wei P, Fiscella M, Baker KP:

DR WPI: 2001-308779/32.

DR N-PSDB: AAD05245.

PT New nucleic acid encoding one of 21 human secreted proteins for  
 PT diagnosing, preventing, treating or ameliorating medical conditions,  
 PT such as autoimmune disease and cancer, and used as a food additive or  
 PT preservative -

PS Claim 11: Page 446; 490pp; English.

CC AAD05220-AAD05282 represent cDNAs corresponding to 21 human secreted  
 CC protein genes, and AAE01352-AAE01413 represent the proteins they encode.  
 CC AAE01415-AAE01433 represent human secreted protein fragments or variants.  
 CC The secreted proteins and their genes are useful for preventing, treating  
 CC or ameliorating medical conditions, e.g., by protein or gene therapy.  
 CC Pathological conditions can be diagnosed by determining the amount of the  
 CC new protein in a sample or by determining the presence of mutations in  
 CC the new genes. Specific uses are described for each of the 21 genes,  
 CC based on the tissues in which they are most highly expressed, and include  
 CC developing products for the diagnosis or treatment of proliferative  
 CC disorders, cancer, tumours, foetal and developmental abnormalities,  
 CC haematopoietic disorders, diseases of the immune system, AIDS, autoimmune  
 CC diseases (e.g., rheumatoid arthritis), inflammation, allergies,  
 CC neurological disorders (e.g., Alzheimer's disease, Parkinson's disease),  
 CC cognitive disorders, schizophrenia, asthma, skin disorders (e.g.,  
 CC psoriasis), sepsis, diabetes, atherosclerosis, cardiovascular disorders,  
 CC angiogenic disorders, kidney disorders, gastrointestinal disorders,  
 CC pregnancy-related disorders, endocrine disorders, and infections. The  
 CC proteins can also be used to aid wound healing and epithelial cell  
 CC proliferation, to prevent skin aging due to sunburn, to maintain organs  
 CC before transplantation, for supporting cell culture of primary tissues,  
 CC to regenerate tissues, to identify their cognate ligands or binding  
 CC partners, and in chemotaxis, and can be used as a food additive or  
 CC preservative to modify storage properties. Antibodies specific for a  
 CC protein of the invention can be used in alleviating symptoms associated  
 CC with the disorders mentioned above, and in diagnostic immunoassays e.g.,  
 CC radioimmunoassay or enzyme linked immunosorbent assay (ELISA).  
 CC The present sequence represents a human secreted protein of  
 CC the invention.

CC Sequence 61 AA:

Alignment Scores:  
 Pred. No.: 117 Length: 61  
 Score: 8.00 Matches: 8  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 1.68% Indels: 0  
 DB: 22 Gaps: 0

US-10-020-441-1 (1-1477) x AAE01377 (1-61)

Oy 235 TGTGCTGCGTTGTGTTGTGA 258

Db 5 CysAlaCysValCysValCysVal 12

RESULT 17

ABG64084

ID ABG64084 standard; Protein: 61 AA.



PR 18-JUN-1999; 99US-0139461.  
PR 18-JUN-1999; 99US-0139462.  
PR 18-JUN-1999; 99US-0139463.  
PR 18-JUN-1999; 99US-0139750.  
PR 18-JUN-1999; 99US-0139753.  
PR 21-JUN-1999; 99US-0139817.  
PR 22-JUN-1999; 99US-0139899.  
PR 23-JUN-1999; 99US-0140353.  
PR 24-JUN-1999; 99US-0140354.  
PR 24-JUN-1999; 99US-0140695.  
PR 28-JUN-1999; 99US-0140823.  
PR 29-JUN-1999; 99US-0140991.  
PR 30-JUN-1999; 99US-0141287.  
PR 01-JUL-1999; 99US-0141842.  
PR 01-JUL-1999; 99US-0142154.  
PR 02-JUL-1999; 99US-0142055.  
PR 06-JUL-1999; 99US-0142350.  
PR 08-JUL-1999; 99US-0142803.  
PR 09-JUL-1999; 99US-0142920.  
PR 12-JUL-1999; 99US-0142977.  
PR 13-JUL-1999; 99US-0143542.  
PR 14-JUL-1999; 99US-0143624.  
PR 15-JUL-1999; 99US-0144005.  
PR 16-JUL-1999; 99US-0144085.  
PR 19-JUL-1999; 99US-0144325.  
PR 19-JUL-1999; 99US-0144331.  
PR 19-JUL-1999; 99US-0144332.  
PR 19-JUL-1999; 99US-0144333.  
PR 19-JUL-1999; 99US-0144334.  
PR 19-JUL-1999; 99US-0144335.  
PR 20-JUL-1999; 99US-0144352.  
PR 20-JUL-1999; 99US-0144632.  
PR 20-JUL-1999; 99US-0144684.  
PR 21-JUL-1999; 99US-0144814.  
PR 21-JUL-1999; 99US-0145086.  
PR 21-JUL-1999; 99US-0145088.  
PR 22-JUL-1999; 99US-0145085.  
PR 22-JUL-1999; 99US-0145087.  
PR 22-JUL-1999; 99US-0145089.  
PR 22-JUL-1999; 99US-0145192.  
PR 23-JUL-1999; 99US-0145145.  
PR 23-JUL-1999; 99US-0145218.  
PR 23-JUL-1999; 99US-0145224.  
PR 26-JUL-1999; 99US-0145276.  
PR 27-JUL-1999; 99US-0145913.  
PR 27-JUL-1999; 99US-0145918.  
PR 27-JUL-1999; 99US-0145919.  
PR 28-JUL-1999; 99US-0145951.  
PR 02-AUG-1999; 99US-0146386.  
PR 02-AUG-1999; 99US-0146388.  
PR 02-AUG-1999; 99US-0146389.  
PR 03-AUG-1999; 99US-0147038.  
PR 04-AUG-1999; 99US-0147204.  
PR 04-AUG-1999; 99US-0147302.  
PR 05-AUG-1999; 99US-0147192.  
PR 06-AUG-1999; 99US-0147260.  
PR 06-AUG-1999; 99US-0147303.  
PR 06-AUG-1999; 99US-0147416.  
PR 09-AUG-1999; 99US-0147493.  
PR 09-AUG-1999; 99US-0147935.  
PR 10-AUG-1999; 99US-0148171.  
PR 11-AUG-1999; 99US-0148319.  
PR 12-AUG-1999; 99US-0148341.  
PR 13-AUG-1999; 99US-0148565.  
PR 13-AUG-1999; 99US-0148684.  
PR 16-AUG-1999; 99US-0149368.  
PR 17-AUG-1999; 99US-0149175.  
PR 18-AUG-1999; 99US-0149426.  
PR 20-AUG-1999; 99US-0149722.  
PR 20-AUG-1999; 99US-0149723.  
PR 20-AUG-1999; 99US-0149929.  
PR 23-AUG-1999; 99US-0149902.

PR 23-AUG-1999; 99US-0149930.  
PR 25-AUG-1999; 99US-0150566.  
PR 26-AUG-1999; 99US-0150864.  
PR 27-AUG-1999; 99US-0151065.  
PR 27-AUG-1999; 99US-0151066.  
PR 27-AUG-1999; 99US-0151080.  
PR 30-AUG-1999; 99US-0151303.  
PR 31-AUG-1999; 99US-0151438.  
PR 01-SEP-1999; 99US-0151930.  
PR 07-SEP-1999; 99US-0152363.  
PR 10-SEP-1999; 99US-0153070.  
PR 13-SEP-1999; 99US-0153758.  
PR 15-SEP-1999; 99US-0154018.  
PR 16-SEP-1999; 99US-0154039.  
PR 20-SEP-1999; 99US-0154779.  
PR 22-SEP-1999; 99US-0155139.  
PR 23-SEP-1999; 99US-0155486.  
PR 24-SEP-1999; 99US-0156458.  
PR 28-SEP-1999; 99US-0156458.  
PR 29-SEP-1999; 99US-0156596.  
PR 04-OCT-1999; 99US-0157117.  
PR 05-OCT-1999; 99US-0157753.  
PR 06-OCT-1999; 99US-0157865.  
PR 07-OCT-1999; 99US-0158029.  
PR 08-OCT-1999; 99US-0158237.  
PR 12-OCT-1999; 99US-0158369.  
PR 13-OCT-1999; 99US-0159293.  
PR 13-OCT-1999; 99US-0159294.  
PR 13-OCT-1999; 99US-0159299.  
PR 14-OCT-1999; 99US-0159329.  
PR 14-OCT-1999; 99US-0159330.  
PR 14-OCT-1999; 99US-0159331.  
PR 14-OCT-1999; 99US-0159637.  
PR 14-OCT-1999; 99US-0159638.  
PR 18-OCT-1999; 99US-0159584.  
PR 21-OCT-1999; 99US-0160741.  
PR 21-OCT-1999; 99US-0160767.  
PR 21-OCT-1999; 99US-0160768.  
PR 21-OCT-1999; 99US-0160770.  
PR 21-OCT-1999; 99US-0160814.  
PR 21-OCT-1999; 99US-0160815.  
PR 22-OCT-1999; 99US-0160980.  
PR 22-OCT-1999; 99US-0160981.  
PR 22-OCT-1999; 99US-0160989.  
PR 25-OCT-1999; 99US-0161404.  
PR 25-OCT-1999; 99US-0161405.  
PR 25-OCT-1999; 99US-0161406.  
PR 26-OCT-1999; 99US-0161359.  
PR 26-OCT-1999; 99US-0161360.  
PR 26-OCT-1999; 99US-0161361.  
PR 28-OCT-1999; 99US-0161920.  
PR 28-OCT-1999; 99US-0161992.  
PR 28-OCT-1999; 99US-0161993.  
PR 29-OCT-1999; 99US-0162142.

## Alignment Scores:

Pred. No.:	116	Length:	67
Score:	8.00	Matches:	8
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	1.68%	Indels:	0
DB:	21	Gaps:	0

US-10-020-441-1 (1-1477) x AAG58997 (1-67)

QY 1143 TTTTGTTTGTTCCTGTGTTT 1166  
|||||

Db 41 PhecysPheValProLeuCysPhe 48

RESULT 19  
AA000781  
ID AA000781 standard; Protein: 79 AA.  
XX

AC AAO00781.  
XX  
DE 06-NOV-2001 (first entry)  
XX  
DE Human polypeptide SEQ ID NO 14673.  
XX  
XX Human: cytokine; cell proliferation; cell differentiation; gene therapy;  
KM vaccine; peptide therapy; stem cell growth factor; haematopoiesis;  
KW tissue growth factor; immunomodulatory; cancer; leukaemia;  
XX nervous system disorders; arthritis; inflammation.  
XX  
OS Homo sapiens.  
XX  
PN WO200164835-A2.  
XX  
PD 07-SEP-2001.  
XX  
PF 26-FEB-2001; 2001WO-US04927.  
XX  
PR 28-FEB-2000; 2000US-0515126.  
PR 18-MAY-2000; 2000US-0577409.  
XX  
PA (HYSE-) HYSEQ INC.  
XX  
PI Tang YT, Liu C, Drmanac RT;  
XX  
DR WPI; 2001-514838/56.  
DR N-PSDB; AAI80712.  
XX  
XX Isolated nucleic acids and polypeptides, useful for preventing  
PT diagnosing and treating e.g. leukaemia, inflammation and immune  
PT disorders -  
XX  
PS Claim 20; SEQ ID NO 14673; 1399pp + Sequence Listing; English.  
XX  
XX The invention relates to human polynucleotides (AAI79941-AAI93841) and  
CC the encoded proteins (AAO00010-AAO13910) that exhibit activity elating to  
CC cytokine, cell proliferation or cell differentiation or which may induce  
CC production of other cytokines in other cell populations. The  
CC polynucleotides and polypeptides are useful in gene therapy, vaccines or  
CC peptide therapy. The polypeptides have various cytokine-like activities,  
CC e.g. stem cell growth factor activity, haematopoiesis regulating  
CC activity, tissue growth factor activity, immunomodulatory activity and  
CC activin/inhibin activity and may be useful in the diagnosis and/or  
CC treatment of cancer, leukaemia, nervous system disorders, arthritis and  
CC inflammation.  
CC Note: The sequence data for this patent did not form part of the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pcl\_sequences.  
XX  
SQ Sequence 79 AA;  
XX  
XX Alignment Scores:  
Pred. No.: 114 Length: 79  
Score: 8.00 Matches: 8  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 1.68% Indels: 0  
DB: 22 Gaps: 0  
XX  
US-10-020-441-1 (1-1477) x AAO00781 (1-79)  
OY 238 GCGTGGCTTGTGTTGTGATGC 261  
DB 14 AlAcYsValCysValCysValCys 21  
XX  
RESULT 20  
AAO08984  
ID AAO08984 standard; Protein: 82 AA.  
XX  
AC AAO08984;  
XX  
XX 06-NOV-2001 (first entry)  
XX

XX  
DE Human polypeptide SEQ ID NO 22876.  
XX  
XX Human: cytokine; cell proliferation; cell differentiation; gene therapy;  
KM vaccine; peptide therapy; stem cell growth factor; haematopoiesis;  
KW tissue growth factor; immunomodulatory; cancer; leukaemia;  
XX nervous system disorders; arthritis; inflammation.  
XX  
OS Homo sapiens.  
XX  
PN WO200164835-A2.  
XX  
PD 07-SEP-2001.  
XX  
PF 26-FEB-2001; 2001WO-US04927.  
XX  
PR 28-FEB-2000; 2000US-0515126.  
PR 18-MAY-2000; 2000US-0577409.  
XX  
PA (HYSE-) HYSEQ INC.  
XX  
PI Tang YT, Liu C, Drmanac RT;  
XX  
DR WPI; 2001-514838/56.  
DR N-PSDB; AAI88915.  
XX  
XX Isolated nucleic acids and polypeptides, useful for preventing  
PT diagnosing and treating e.g. leukaemia, inflammation and immune  
PT disorders -  
XX  
PS Claim 20; SEQ ID NO 22876; 1399pp + Sequence Listing; English.  
XX  
XX The invention relates to human polynucleotides (AAI79941-AAI93841) and  
CC the encoded proteins (AAO00010-AAO13910) that exhibit activity elating to  
CC cytokine, cell proliferation or cell differentiation or which may induce  
CC production of other cytokines in other cell populations. The  
CC polynucleotides and polypeptides are useful in gene therapy, vaccines or  
CC peptide therapy. The polypeptides have various cytokine-like activities,  
CC e.g. stem cell growth factor activity, haematopoiesis regulating  
CC activity, tissue growth factor activity, immunomodulatory activity and  
CC activin/inhibin activity and may be useful in the diagnosis and/or  
CC treatment of cancer, leukaemia, nervous system disorders, arthritis and  
CC inflammation.  
CC Note: The sequence data for this patent did not form part of the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pcl\_sequences.  
XX  
SQ Sequence 82 AA;  
XX  
XX Alignment Scores:  
Pred. No.: 114 Length: 82  
Score: 8.00 Matches: 8  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 1.68% Indels: 0  
DB: 22 Gaps: 0  
XX  
US-10-020-441-1 (1-1477) x AAO08984 (1-82)  
OY 691 TGTGTCGTGTATCTGTTGTG 714  
DB 8 CysValCysValSerValCysVal 15  
XX  
RESULT 21  
AAM84230  
ID AAM84230 standard; Protein: 83 AA.  
XX  
AC AAM84230;  
XX  
XX 07-NOV-2001 (first entry)  
XX  
XX Human immune/haematopoietic antigen SEQ ID NO.11823.  
XX

KW Human; immune; haematopoietic; immune/haematopoietic antigen; cancer;  
KM cytostatic; gene therapy; vaccine; metastasis.  
XX  
OS Homo sapiens.  
XX  
PN WO200157182-A2.  
XX  
PD 09-AUG-2001.  
XX  
PF 17-JAN-2001; 2001WO-US01354.  
XX  
PR 31-JAN-2000; 2000US-0179065.  
PR 04-FEB-2000; 2000US-0180628.  
PR 24-FEB-2000; 2000US-0184664.  
PR 02-MAR-2000; 2000US-0186350.  
PR 16-MAR-2000; 2000US-0189874.  
PR 17-MAR-2000; 2000US-0190076.  
PR 18-APR-2000; 2000US-0198123.  
PR 19-MAY-2000; 2000US-0205515.  
PR 07-JUN-2000; 2000US-0209467.  
PR 28-JUN-2000; 2000US-0214886.  
PR 30-JUN-2000; 2000US-0215135.  
PR 07-JUL-2000; 2000US-0216647.  
PR 07-JUL-2000; 2000US-0216880.  
PR 11-JUL-2000; 2000US-0217487.  
PR 11-JUL-2000; 2000US-0217496.  
PR 14-JUL-2000; 2000US-0218290.  
PR 26-JUL-2000; 2000US-0220963.  
PR 26-JUL-2000; 2000US-0220964.  
PR 14-AUG-2000; 2000US-0224518.  
PR 14-AUG-2000; 2000US-0224519.  
PR 14-AUG-2000; 2000US-0225213.  
PR 14-AUG-2000; 2000US-0225214.  
PR 14-AUG-2000; 2000US-0225266.  
PR 14-AUG-2000; 2000US-0225267.  
PR 14-AUG-2000; 2000US-0225268.  
PR 14-AUG-2000; 2000US-0225270.  
PR 14-AUG-2000; 2000US-0225447.  
PR 14-AUG-2000; 2000US-0225757.  
PR 14-AUG-2000; 2000US-0225758.  
PR 14-AUG-2000; 2000US-0225759.  
PR 18-AUG-2000; 2000US-0226279.  
PR 22-AUG-2000; 2000US-0226681.  
PR 22-AUG-2000; 2000US-0226868.  
PR 22-AUG-2000; 2000US-0227182.  
PR 23-AUG-2000; 2000US-0227009.  
PR 30-AUG-2000; 2000US-0228924.  
PR 01-SEP-2000; 2000US-0229287.  
PR 01-SEP-2000; 2000US-0229343.  
PR 01-SEP-2000; 2000US-0229344.  
PR 01-SEP-2000; 2000US-0229345.  
PR 05-SEP-2000; 2000US-0229509.  
PR 05-SEP-2000; 2000US-0229513.  
PR 06-SEP-2000; 2000US-0230437.  
PR 06-SEP-2000; 2000US-0230438.  
PR 08-SEP-2000; 2000US-0231242.  
PR 08-SEP-2000; 2000US-0231243.  
PR 08-SEP-2000; 2000US-0231244.  
PR 08-SEP-2000; 2000US-0231413.  
PR 08-SEP-2000; 2000US-0231414.  
PR 08-SEP-2000; 2000US-0232080.  
PR 08-SEP-2000; 2000US-0232081.  
PR 12-SEP-2000; 2000US-0231968.  
PR 14-SEP-2000; 2000US-0232397.  
PR 14-SEP-2000; 2000US-0232398.  
PR 14-SEP-2000; 2000US-0232399.  
PR 14-SEP-2000; 2000US-0232400.  
PR 14-SEP-2000; 2000US-0232401.  
PR 14-SEP-2000; 2000US-0233063.  
PR 14-SEP-2000; 2000US-0233064.  
PR 14-SEP-2000; 2000US-0233065.  
PR 21-SEP-2000; 2000US-0234223.  
PR 21-SEP-2000; 2000US-0234274.  
  
PR 25-SEP-2000; 2000US-0234997.  
PR 25-SEP-2000; 2000US-0234998.  
PR 26-SEP-2000; 2000US-0235484.  
PR 27-SEP-2000; 2000US-0235834.  
PR 27-SEP-2000; 2000US-0235836.  
PR 29-SEP-2000; 2000US-0236327.  
PR 29-SEP-2000; 2000US-0236367.  
PR 29-SEP-2000; 2000US-0236368.  
PR 29-SEP-2000; 2000US-0236369.  
PR 29-SEP-2000; 2000US-0236370.  
PR 02-OCT-2000; 2000US-0236802.  
PR 02-OCT-2000; 2000US-0237037.  
PR 02-OCT-2000; 2000US-0237038.  
PR 02-OCT-2000; 2000US-0237039.  
PR 02-OCT-2000; 2000US-0237040.  
PR 13-OCT-2000; 2000US-0239935.  
PR 13-OCT-2000; 2000US-0239937.  
PR 20-OCT-2000; 2000US-0240960.  
PR 20-OCT-2000; 2000US-0241221.  
PR 20-OCT-2000; 2000US-0241785.  
PR 20-OCT-2000; 2000US-0241786.  
PR 20-OCT-2000; 2000US-0241787.  
PR 20-OCT-2000; 2000US-0241808.  
PR 20-OCT-2000; 2000US-0241809.  
PR 20-OCT-2000; 2000US-0241826.  
PR 01-NOV-2000; 2000US-0244617.  
PR 08-NOV-2000; 2000US-0246474.  
PR 08-NOV-2000; 2000US-0246475.  
PR 08-NOV-2000; 2000US-0246476.  
PR 08-NOV-2000; 2000US-0246477.  
PR 08-NOV-2000; 2000US-0246478.  
PR 08-NOV-2000; 2000US-0246523.  
PR 08-NOV-2000; 2000US-0246524.  
PR 08-NOV-2000; 2000US-0246525.  
PR 08-NOV-2000; 2000US-0246526.  
PR 08-NOV-2000; 2000US-0246527.  
PR 08-NOV-2000; 2000US-0246528.  
PR 08-NOV-2000; 2000US-0246532.  
PR 08-NOV-2000; 2000US-0246609.  
PR 08-NOV-2000; 2000US-0246610.  
PR 08-NOV-2000; 2000US-0246611.  
PR 08-NOV-2000; 2000US-0246613.  
PR 17-NOV-2000; 2000US-0249207.  
PR 17-NOV-2000; 2000US-0249208.  
PR 17-NOV-2000; 2000US-0249209.  
PR 17-NOV-2000; 2000US-0249210.  
PR 17-NOV-2000; 2000US-0249211.  
PR 17-NOV-2000; 2000US-0249212.  
PR 17-NOV-2000; 2000US-0249213.  
PR 17-NOV-2000; 2000US-0249214.  
PR 17-NOV-2000; 2000US-0249215.  
PR 17-NOV-2000; 2000US-0249216.  
PR 17-NOV-2000; 2000US-0249217.  
PR 17-NOV-2000; 2000US-0249218.  
PR 17-NOV-2000; 2000US-0249244.  
PR 17-NOV-2000; 2000US-0249245.  
PR 17-NOV-2000; 2000US-0249264.  
PR 17-NOV-2000; 2000US-0249265.  
PR 17-NOV-2000; 2000US-0249267.  
PR 17-NOV-2000; 2000US-0249299.  
PR 17-NOV-2000; 2000US-0249300.  
PR 01-DEC-2000; 2000US-0250160.  
PR 01-DEC-2000; 2000US-0250391.  
PR 05-DEC-2000; 2000US-0251030.  
PR 05-DEC-2000; 2000US-0251988.  
PR 05-DEC-2000; 2000US-0256719.  
PR 06-DEC-2000; 2000US-0251479.  
PR 08-DEC-2000; 2000US-0251856.  
PR 08-DEC-2000; 2000US-0251868.  
PR 08-DEC-2000; 2000US-0251869.  
PR 08-DEC-2000; 2000US-0251989.  
PR 08-DEC-2000; 2000US-0251990.  
PR 11-DEC-2000; 2000US-0254097.

PR 05-JAN-2001: 2001US-0259678.  
XX  
XX (HUMA-) HUMAN GENOME SCI INC.  
XX  
XX Rosen CA, Barash SC, Ruben SM;  
XX  
XX WPI: 2001-483426/52.  
DR N-PSDB: AAK57011.  
PT Nucleic acids encoding human immune/hematopoietic antigen polypeptides,  
PT useful for preventing, diagnosing and/or treating cancers and  
PT metastasis -  
XX  
PS Claim 11: SEQ ID NO 11823: 3071pp + Sequence Listing; English.  
XX  
XX AAK54951 to AAK64702 encode the human immune/hematopoietic antigen (I)  
CC amino acid sequences given in AAM82170 to AAM91921. (I) have cytostatic  
CC activity, and can be used in gene therapy and vaccine production. (I)  
CC proteins and polynucleotides may be used in the prevention, diagnosis and  
CC treatment of diseases associated with inappropriate (I) expression. For  
CC example, they may be used to treat disorders associated with decreased  
CC expression by rectifying mutations or deletions in a patient's genome  
CC that affect the activity of (I) by expressing inactive proteins or to  
CC supplement the patients own production of (I). Additionally, (I)  
CC polynucleotides may be used to produce the secreted (I), by inserting  
CC the nucleic acids into a host cell and culturing the cell to express the  
CC protein. (I) proteins and polynucleotides may be used to prevent,  
CC diagnose and treat immune/hematopoietic-related diseases, especially  
CC cancers and cancer metastases of hematopoietic-derived cells. AAK64703  
CC to AAK87694 represent human immune/hematopoietic antigen genomic  
CC sequences from the present invention. AAK54942 to AAK54950 and AAM82169  
CC represent sequences used in the exemplification of the present invention.  
XX  
XX SQ Sequence 83 AA:  
SQ  
Alignment Scores:  
Pred. No.: 114 Length: 83  
Score: 8.00 Matches: 8  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 1.68% Indels: 0  
Gaps: 0  
US-10-020-441-1 (1-1477) x AAM84230 (1-83)  
QY 35 TGCACACCGCACTGAATTCCCGT 58  
DB 37 CysAsnThrAlaLeuAsnSerArg 44  
RESULT 22  
AAB43279  
ID AAB43279 standard; Protein: 84 AA.  
XX  
XX AAB43279:  
AC  
XX  
XX 08-FEB-2001 (first entry)  
DT  
XX  
DE Human ORFX ORF3043 polypeptide sequence SEQ ID NO:6086.  
XX  
XX Human: open reading frame; ORFX: detection; cytostatic; hepatotropic;  
KM vulnery; antiprosclatic; antiparkinsonian; nootropic; neuroprotective;  
KM anticonvulsant; osteopathic; antiarthritic; immunosuppressant; cardiant;  
KM immunostimulant; thrombolytic; coagulant; vasotropic; antidiabetic;  
KM hypotensive; dermatological; immunosuppressive; antiinflammatory;  
KM antiviral; antibacterial; antifungal; antirheumatic; antithyroid;  
KM antianemic; gene therapy; cancer; proliferative disorder; hypertension;  
KM neurodegenerative disorder; osteoarthritis; graft vs host disease;  
KM cardiovascular disease; diabetes mellitus; hypothyroidism; SCID; AIDS;  
KM cholesterol ester storage; systemic lupus erythematosus; infection;  
KM severe combined immunodeficiency; malaria; autoimmune disorder; asthma;  
KM allergy; aplastic anaemia; nocturnal haemoglobinuria; burn; wound;  
KM bone damage; cartilage damage; antiinflammatory disease; coagulation;  
KM thrombosis; contraceptive.

XX  
XX Homo sapiens.  
OS  
XX  
XX WO200058473-A2.  
PN  
XX  
XX 05-OCT-2000.  
PD  
XX  
XX 31-MAR-2000: 2000WO-US08621.  
PE  
XX  
XX 31-MAR-1999: 99US-0127607.  
PR 02-APR-1999: 99US-0127636.  
PR 05-APR-1999: 99US-0127728.  
PR 30-MAR-2000: 2000US-0540763.  
XX  
XX (CURA-) CURAGEN CORP.  
PA  
XX  
XX Shinkets RA, Leach M;  
PI  
XX  
XX WPI: 2000-602362/57.  
DR N-PSDB: AAC77486.  
DR  
XX  
XX Novel nucleic acids and peptides derived from open reading frame X,  
PT useful for treating e.g. cancers, proliferative disorders,  
PT neurodegenerative disorders and cardiovascular disease -  
XX  
XX Claim 11: Page 5268: 5507pp; English.  
PS  
XX  
XX AAC74446 to AAC77606 encode the proteins given in AAB40237 to AAB43397,  
CC which represent the human ORFX open reading frames 1 to 3161. The ORFX  
CC sequences have activities such as: cytostatic; hepatotropic; vulnery;  
CC antiprosclatic; antiparkinsonian; nootropic; neuroprotective;  
CC osteopathic; anticonvulsant; antiarthritic; immunosuppressant;  
CC immunostimulant; cardiant; thrombolytic; coagulant; vasotropic;  
CC antidiabetic; hypotensive; dermatological; immunosuppressive;  
CC antiinflammatory; antibacterial; antiviral; antifungal; antirheumatic;  
CC antithyroid; and antianemic. The sequences can be used for determining  
CC the presence of or predisposition to, or preventing or treating  
CC pathological conditions associated with an ORFX-associated disorder. The  
CC nucleic acids can be used to express ORFX proteins in gene therapy  
CC vectors. The proteins and nucleic acids may be used to treat cancers,  
CC proliferative disorders, neurodegenerative disorders, osteoarthritis,  
CC graft vs host disease, cardiovascular disease, diabetes mellitus,  
CC hypertension, hypothyroidism, cholesterol ester storage, systemic lupus  
CC erythematosus, severe combined immunodeficiency (SCID), AIDS, viral,  
CC bacterial or fungal infection, malaria, autoimmune disorders, asthma,  
CC allergies, aplastic anaemia, burns, wounds, bone and cartilage damage,  
CC nocturnal haemoglobinuria, antiinflammatory disease; to enhance  
CC coagulation; to inhibit thrombosis; and as a contraceptive.  
XX  
XX SQ Sequence 84 AA:  
SQ  
Alignment Scores:  
Pred. No.: 114 Length: 84  
Score: 8.00 Matches: 8  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 1.68% Indels: 0  
Gaps: 0  
US-10-020-441-1 (1-1477) x AAB43279 (1-84)  
QY 235 TGTCCGTGCGTTGTGTTGTGCTA 258  
DB 46 CysAlaCysValCysValCysVal 53  
RESULT 23  
ABP33051  
ID ABP33051 standard; Protein: 90 AA.  
XX  
XX AAB33051:  
AC  
XX  
XX 09-JUL-2002 (first entry)  
DT  
XX

DE Human ORF2024 protein, SEQ ID NO:4048.  
 XX  
 KW Human: ORF: open reading frame; ORFX: drug screening; diagnosis;  
 KW disease monitoring; cytokine; cell proliferation; cell differentiation;  
 KW immune modulation; haematopoiesis regulation; tissue growth;  
 KW angiogenesis; activin; inhibin; chemotactic; chemokine; haemostatic;  
 KW thrombolytic; tumour inhibition; bodily characteristics; fertility;  
 KW behaviour; cancer; proliferative disorder; neurological disorder;  
 KW cardiovascular disease; immune system disorder; organ transplantation;  
 KW tissue growth disorder; tissue regeneration disorder; diabetes mellitus;  
 KW hypothyroidism; cholesterol ester storage disease; infection; vulnerability;  
 KW vasoprotective; antipsoriatic; antidiabetic; cytosolic; neutropenic;  
 KW neuroprotective; antiatherosclerotic; anticoagulant; thrombolytic;  
 KW cardiant; hypotensive; antihypertoid; antiinflammatory; immunomodulator;  
 KW dermatological; analgesic; virucide; antibacterial; fungicide.  
 KW  
 XX Homo sapiens.  
 OS  
 PN WO200190366-A2.  
 XX  
 PD 29-NOV-2001.  
 XX  
 PF 24-MAY-2001; 2001WO-US17076.  
 XX  
 PR 24-MAY-2000; 2000US-206690P.  
 XX  
 PA (CURA-) CURAGEN CORP.  
 XX  
 PI Leach MD, Shinkets RA;  
 XX  
 DR WPI: 2002-106200/14.  
 DR N-PSDB: ABN77077.  
 XX  
 PT Novel human polypeptides and polynucleotides useful for diagnosing,  
 PT preventing and treating cardiovascular disease, neurodegenerative,  
 PT hyperproliferative disorders and disorders related to organ  
 PT transplantation -  
 PT  
 PS Claim 10; Page 1269; 2508pp; English.  
 PS  
 XX Sequences ABP31028-ABP35561 represent 4534 novel human proteins  
 CC designated ORF (open reading frame) 1-4534, and sequences ABN75054-  
 CC ABN75587 represent cDNAs encoding them. The invention also encompasses  
 CC polypeptides at least 80% identical to the ORF1-ORF4534 (collectively  
 CC referred to as ORFX) proteins, polynucleotides at least 85% identical to  
 CC the ORFX nucleic acid sequences, vectors and host cells comprising ORFX  
 CC polynucleotides, the recombinant production of ORFX proteins, antibodies  
 CC specific for ORFX proteins, methods of detecting ORFX polynucleotides and  
 CC polypeptides, methods of screening for modulators of ORFX expression or  
 CC activity, and methods of screening individuals for a predisposition to an  
 CC ORFX-associated disorder. The ORFX proteins of the invention have a wide  
 CC range of biological activities, such as cytokine, cell proliferation,  
 CC cell differentiation, immune modulation, haematopoiesis regulation,  
 CC tissue growth, angiogenesis, activin or inhibin activity, chemotactic/  
 CC chemokinetic activity, haemostatic activity, thrombolytic activity,  
 CC receptor/ligand, antiinflammatory activity, tumour inhibition activity,  
 CC and antiinfective activity, and may also be involved in the determination  
 CC of bodily characteristics, fertility and behaviour. ORFX proteins,  
 CC nucleic acids and antibodies may be used in the treatment of cancers,  
 CC other proliferative disorders such as psoriasis and benign tumours,  
 CC neurological disorders such as epilepsy and Alzheimer's disease,  
 CC cardiovascular diseases, immune system disorders, disorders related to  
 CC organ transplantation, disorders of tissue growth and regeneration,  
 CC diseases such as diabetes mellitus, hypothyroidism, and cholesterol ester  
 CC storage disease, and infectious diseases caused by viral, bacterial,  
 CC fungal and other pathogens. ORFX nucleic acids may also be used as a  
 CC source of primers and probes, in the detection of ORFX genomic sequences  
 CC or transcripts, in the identification and cloning of homologous  
 CC sequences, in genetic diagnosis, and in forensic biology. The ORFX  
 CC nucleic acids may additionally be used to produce transgenic animals  
 CC which may be useful for studying the function and/or activity of ORFX  
 CC protein, and in drug screening. The ORFX proteins may also be used as  
 CC immunogens to generate specific antibodies, which are useful in the

CC diagnosis, treatment and monitoring of ORFX-associated diseases.  
 XX  
 SQ Sequence 90 AA:  
 XX  
 Alignment Scores:  
 Pred. No.: 113 Length: 90  
 Score: 8.00 Matches: 8  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 1.68% Indels: 0  
 DB: 23 Gaps: 0  
 US-10-020-441-1 (1-1477) x ABP33051 (1-90)  
 QY 232 GCGTGGCGTGGCTTGTGTTGT 255  
 Db 11 AlacysAlacysValcysValcys 18  
 |||||  
 RESULT 24  
 AA002862  
 ID AA002862 standard; Protein: 92 AA.  
 XX  
 AC AA002862;  
 XX  
 DT 06-NOV-2001 (first entry)  
 XX  
 DE Human polypeptide SEQ ID NO 16754.  
 XX  
 KW Human: cytokine; cell proliferation; cell differentiation; gene therapy;  
 KW vaccine; peptide therapy; stem cell growth factor; haematopoiesis;  
 KW tissue growth factor; immunomodulatory; cancer; leukaemia;  
 KW nervous system disorders; arthritis; inflammation.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO200164835-A2.  
 XX  
 PD 07-SEP-2001.  
 XX  
 PF 26-FEB-2001; 2001WO-US04927.  
 XX  
 PR 28-FEB-2000; 2000US-0515126.  
 PR 18-MAY-2000; 2000US-0577409.  
 XX  
 PA (HYSE-) HYSEQ INC.  
 XX  
 PI Tang YT, Liu C, Drmanac RT;  
 XX  
 DR WPI: 2001-514838/56.  
 DR N-PSDB: AAI82793.  
 XX  
 PT Isolated nucleic acids and polypeptides, useful for preventing  
 PT diagnosing and treating e.g. leukaemia, inflammation and immune  
 PT disorders -  
 PT  
 PS Claim 20; SEQ ID NO 16754; 1399pp + sequence listing; English.  
 PS  
 XX The invention relates to human polynucleotides (AA179941-AA193841) and  
 CC the encoded proteins (AA000010-AA013910) that exhibit activity elating to  
 CC cytokine, cell proliferation or cell differentiation or which may induce  
 CC production of other cytokines in other cell populations. The  
 CC polynucleotides and polypeptides are useful in gene therapy, vaccines or  
 CC peptide therapy. The polypeptides have various cytokine-like activities,  
 CC e.g. stem cell growth factor activity, haematopoiesis regulating  
 CC activity, tissue growth factor activity, immunomodulatory activity and  
 CC activin/inhibin activity and may be useful in the diagnosis and/or  
 CC treatment of cancer, leukaemia, nervous system disorders, arthritis and  
 CC inflammation.  
 CC Note: The sequence data for this patent did not form part of the printed  
 CC specification, but was obtained in electronic format directly from WIPO  
 CC at ftp.wipo.int/pub/published\_pct\_sequences.  
 CC  
 XX Sequence 92 AA:

Alignment Scores:

Pred. No.:	113	Length:	92
Score:	8.00	Matches:	8
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	1.68%	Indels:	0
DB:	22	Gaps:	0

US-10-020-441-1 (1-1477) x AA002862 (1-92)

OY 691 TGTGTGTCGTACTGCTTGTGTCG 714  
|||||

Db 72 CysValCysValSerValCysVal 79

RESULT 25  
AAB41641

ID AAB41641 standard; Protein: 102 AA.

AC AAB41641:

XX 08-FEB-2001 (first entry)

DE Human ORFX ORF1405 polypeptide sequence SEQ ID NO:2810.

XX Human: open reading frame; ORFX: detection; cytostatic; hepatotropic; vulnery; antiproliferative; antiparkinsonian; neurotrophic; neuroprotective; anticonvulsant; osteopathic; antiarthritic; immunosuppressant; cardiant; immunostimulant; thrombolytic; coagulant; vasotropic; antidiabetic; hypotensive; dermatological; immunosuppressive; antirheumatic; antihypertensive; antiviral; antibacterial; antifungal; antirheumatic; antihypertensive; antineoplastic; gene therapy; cancer; proliferative disorder; hypertension; neurodegenerative disorder; osteoarthritis; graft vs host disease; cardiovascular disease; diabetes mellitus; hypothyroidism; SCID; AIDS; cholesterol ester storage; systemic lupus erythematosus; infection; severe combined immunodeficiency; malaria; autoimmune disorder; asthma; allergy; aplastic anaemia; nocturnal haemoglobinuria; burn; wound; bone damage; cartilage damage; antiinflammatory disease; coagulation; thrombosis; contraceptive.

XX Homo sapiens.

OS WO200058473-A2.

PN 05-OCT-2000.

XX 31-MAR-2000; 2000WO-US08621.

XX 31-MAR-1999; 99US-0127607.

PR 02-APR-1999; 99US-0127636.

PR 05-APR-1999; 99US-0127728.

PR 30-MAR-2000; 2000US-0540763.

XX (CURA-) CURAGEN CORP.

XX Slimkets RA, Leach M;

PI WPI; 2000-602362/57.

XX N-PSDB; AAC75850.

DR Novel nucleic acids and peptides derived from open reading frame X, useful for treating e.g. cancers, proliferative disorders, neurodegenerative disorders and cardiovascular disease -

XX Claim 11; Page 2049; 5507pp; English.

XX AAC74446 to AAC77606 encode the proteins given in AAB40237 to AAB43397, CC sequences represent the human ORFX open reading frames 1 to 3161. The ORFX CC sequences have activities such as: cytostatic; hepatotropic; vulnery; antiproliferative; antiparkinsonian; neurotrophic; neuroprotective; antidiabetic; anticonvulsant; antiarthritic; immunosuppressant; osteopathic; cardiant; thrombolytic; coagulant; vasotropic; immunostimulant; allergy; aplastic anaemia; nocturnal haemoglobinuria; burn; wound; bone damage; cartilage damage; antiinflammatory disease; coagulation; thrombosis; contraceptive; immunosuppressive;

CC antiinflammatory; antibacterial; antiviral; antifungal; antirheumatic; CC antihypertensive; and antineoplastic. The sequences can be used for determining CC the presence of or predisposition to, or preventing or treating CC pathological conditions associated with an ORFX-associated disorder. The CC nucleic acids can be used to express ORFX proteins in gene therapy CC vectors. The proteins and nucleic acids may be used to treat cancers, CC proliferative disorders, neurodegenerative disorders, osteoarthritis, CC graft vs host disease, cardiovascular disease, diabetes mellitus, CC hypertension, hypothyroidism, cholesterol ester storage, systemic lupus CC erythematosus, severe combined immunodeficiency (SCID), AIDS, viral, CC bacterial or fungal infection, malaria, autoimmune disorders, asthma, CC allergies, aplastic anaemia, burns, wounds, bone and cartilage damage, CC nocturnal haemoglobinuria, antiinflammatory disease; to enhance CC coagulation; to inhibit thrombosis; and as a contraceptive.

XX Sequence 102 AA;

Alignment Scores:

Pred. No.:	112	Length:	102
Score:	8.00	Matches:	8
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	1.68%	Indels:	0
DB:	21	Gaps:	0

US-10-020-441-1 (1-1477) x AAB41641 (1-102)

OY 228 GTTTCGTCGTGTCGTGTCGT 251  
|||||

Db 7 ValCysValCysValArgLeuCys 14

RESULT 26  
AAU47000

ID AAU47000 standard; Protein: 103 AA.

XX AAU47000:

XX 27-FEB-2002 (first entry)

DE Propionibacterium acnes immunogenic protein #7896.

XX SAPHO syndrome; synovitis; acne; pustulosis; hypertrophy; osteomyelitis; CC uveitis; endophthalmitis; bone; joint; central nervous system; ELISA; CC inflammatory lesion; acne vulgaris; enzyme linked immunosorbent assay; CC dermatological; osteopathic; neuroprotectant.

XX Propionibacterium acnes.

OS WO200181581-A2.

PN 01-NOV-2001.

XX 20-APR-2001; 2001WO-US12865.

XX 21-APR-2000; 2000US-199047P.

PR 02-JUN-2000; 2000US-208841P.

PR 07-JUL-2000; 2000US-216747P.

XX (CORI-) CORIXA CORP.

XX Sleiky YAM, Persing DH, Mitcham JL, Wang SS, Bhatia A;

PI L'maisonneuve J, Zhang Y, Jen S, Carter D;

XX WPI; 2001-616774/71.

DR N-PSDB; AAS59536.

XX Propionibacterium acnes polypeptides and nucleic acids useful for PT vaccinating against and diagnosing infections, especially useful for PT treating acne vulgaris -

XX Example 1; SEQ ID NO 8195; 1069pp; English.

XX Sequences AAU39105-AAU68017 represent Propionibacterium acnes immunogenic



CC polypeptides. The proteins and their associated DNA sequences are used in  
CC the treatment, prevention and diagnosis of medical conditions caused by  
CC P. acnes. The disorders include SAPHO syndrome (synovitis, acne,  
CC pustulosis, hyperostosis and osteomyelitis), uveitis and endophthalmitis.  
CC P. acnes is also involved in infections of bone, joints and the central  
CC nervous system, however it is particularly involved in the inflammatory  
CC lesions associated with acne vulgaris. A method for detecting the  
CC presence or absence of P. acnes in a patient comprises contacting a  
CC sample with a binding agent that binds to the proteins of the invention  
CC and determining the amount of bound protein in the sample. The  
CC polypeptides may be used as antigens in the production of antibodies  
CC specific for P. acnes proteins. These antibodies can be used to  
CC downregulate expression and activity of P. acnes polypeptides and  
CC therefore treat P. acnes infections. The antibodies may also be used as  
CC diagnostic agents for determining P. acnes presence, for example, by  
CC enzyme linked immunosorbent assay (ELISA).  
CC Note: The sequence data for this patent did not form part of the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pct\_sequences.

XX SQ Sequence 103 AA;

Alignment Scores:  
Pred. No.: 111 Length: 103  
Score: 8.00 Matches: 8  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 1.68% Indels: 0  
DB: 22 Gaps: 0

US-10-020-441-1 (1-1477) x AA047000 (1-103)

OY 1360 TCAGGTCCTACTCGGTGCTGCT 1383

DB 75 SerArgSerSerThrArgCysArg 82

RESULT 27

AA004304  
ID AA004304 standard; Protein: 124 AA.

XX AC AA004304;

XX DT 06-NOV-2001 (first entry)

XX DE Human polypeptide SEQ ID NO 18196.

XX KW Human; cytokine; cell proliferation; cell differentiation; gene therapy;

KW vaccine; peptide therapy; stem cell growth factor; haematopoiesis;

KW tissue growth factor; immunomodulatory; cancer; leukaemia;

KW nervous system disorders; arthritis; inflammation.

XX OS Homo sapiens.

XX PN WO200164835-A2.

XX PD 07-SEP-2001.

XX PF 26-FEB-2001; 2001WO-US04927.

XX PR 28-FEB-2000; 2000US-0515126.

XX PR 18-MAY-2000; 2000US-0577409.

XX PA (HYSE-) HYSEQ INC.

XX PI Tang YT, Liu C, Drmanac RT;

XX DR WPI: 2001-514838/56.

XX DR N-PSDB; AAI84235.

XX PT Isolated nucleic acids and polypeptides, useful for preventing  
XX disorders -  
XX diagnosing and treating e.g. leukaemia, inflammation and immune  
XX disorders -

PS Claim 20; SEQ ID NO 18196; 1399pp + Sequence Listing: English.

XX CC The invention relates to human polynucleotides (AAI79941-AAI93841) and  
XX CC the encoded proteins (AA000010-AA013910) that exhibit activity relating to  
XX CC cytokine, cell proliferation or cell differentiation or which may induce  
XX CC production of other cytokines in other cell populations. The  
XX CC polynucleotides and polypeptides are useful in gene therapy, vaccines or  
XX CC peptide therapy. The polypeptides have various cytokine-like activities,  
XX CC e.g. stem cell growth factor activity, haematopoiesis regulating  
XX CC activity, tissue growth factor activity, immunomodulatory activity and  
XX CC activity/inhibin activity and may be useful in the diagnosis and/or  
XX CC treatment of cancer, leukaemia, nervous system disorders, arthritis and  
XX CC inflammation.  
XX CC Note: The sequence data for this patent did not form part of the printed  
XX CC specification, but was obtained in electronic format directly from WIPO  
XX CC at ftp.wipo.int/pub/published\_pct\_sequences.

XX SQ Sequence 124 AA;

Alignment Scores:  
Pred. No.: 110 Length: 124  
Score: 8.00 Matches: 8  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 1.68% Indels: 0  
DB: 22 Gaps: 0

US-10-020-441-1 (1-1477) x AA004304 (1-124)

OY 238 GCGTGCCTTGTGTTGTGTATGC 261

DB 72 AlacysValcysValcysValcys 79

RESULT 28

AA001446  
ID AA001446 standard; Protein: 140 AA.

XX AC AA001446;

XX DT 06-NOV-2001 (first entry)

XX DE Human polypeptide SEQ ID NO 15338.

XX KW Human; cytokine; cell proliferation; cell differentiation; gene therapy;

KW vaccine; peptide therapy; stem cell growth factor; haematopoiesis;

KW tissue growth factor; immunomodulatory; cancer; leukaemia;

KW nervous system disorders; arthritis; inflammation.

XX OS Homo sapiens.

XX PN WO200164835-A2.

XX PD 07-SEP-2001.

XX PF 26-FEB-2001; 2001WO-US04927.

XX PR 28-FEB-2000; 2000US-0515126.

XX PR 18-MAY-2000; 2000US-0577409.

XX PA (HYSE-) HYSEQ INC.

XX PI Tang YT, Liu C, Drmanac RT;

XX DR WPI: 2001-514838/56.

XX DR N-PSDB; AAI81377.

XX PT Isolated nucleic acids and polypeptides, useful for preventing  
XX disorders -  
XX diagnosing and treating e.g. leukaemia, inflammation and immune  
XX disorders -  
XX Claim 20; SEQ ID NO 15338; 1399pp + Sequence Listing: English.

XX The invention relates to human polynucleotides (AAI79941-AAI93841) and

CC the encoded proteins (AA000010-AA013910) that exhibit activity elating to  
CC cytokine, cell proliferation or cell differentiation or which may induce  
CC production of other cytokines in other cell populations. The  
CC polynucleotides and polypeptides are useful in gene therapy, vaccines or  
CC peptide therapy. The polypeptides have various cytokine-like activities,  
CC e.g. stem cell growth factor activity, haematopoiesis regulating  
CC activity, tissue growth factor activity, immunomodulatory activity and  
CC activin/inhibin activity and may be useful in the diagnosis and/or  
CC treatment of cancer, leukaemia, nervous system disorders, arthritis and  
CC inflammation.  
CC Note: The sequence data for this patent did not form part of the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pct\_sequences.  
CC  
SQ Sequence 140 AA:  
  
Alignment Scores:  
Pred. No.: 108 Length: 140  
Score: 8.00 Matches: 8  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 1.68% Indels: 0  
DB: 22 Gaps: 0  
  
US-10-020-441-1 (1-1477) x AA01446 (1-140)  
  
QY 238 GCGTGCCTTGTGTTGTGATGC 261  
ID 118 AAcysValcysValcysValcys 125  
Db 118 AAcysValcysValcysValcys 125  
  
RESULT 29  
ABB57252  
ID ABB57252 standard; Protein: 176 AA.  
AC ABB57252;  
XX  
AC 07-MAR-2002 (first entry)  
XX  
DE Mouse ischaemic condition related protein sequence SEQ ID NO:691.  
XX  
XX  
KM Mouse; Ischaemia; compressive Ischaemia; occlusive ischaemia;  
XX vasospastic Ischaemia; Ischaemic condition; Ischaemic disease.  
XX  
OS Mus musculus.  
XX  
PN WO200188188-A2.  
XX  
PD 22-NOV-2001.  
XX  
PF 18-MAY-2001; 2001WO-JP04192.  
XX  
PR 18-MAY-2000; 2000JP-0145977.  
XX  
PA (UYNI-) UNIV NIHON SCHOOL JURIDICAL PERSON.  
XX  
PI Ishikawa K, Asai S, Takahashi Y, Nagata T, Ishii Y;  
XX WPI; 2002-034733/04.  
XX N-PSDB; ABI99659.  
XX  
PT Examining the ischemic condition (e.g. occlusive ischemia) by measuring  
PT expression levels of particular genes defined in the specification or  
PT by determining the expression profile of a gene group comprising these  
PT genes -  
XX  
XX Claim 2; Page 1721-1722; 2690pp; English.  
XX  
XX The present invention describes a method for examining ischaemic  
CC conditions, comprising measuring the expression levels of particular  
CC genes (1) in a test sample or determining the expression profile of a  
CC gene group in the sample comprising genes selected from (1). The method  
CC is useful for examining the ischaemic condition (e.g. compressive  
CC ischaemia, occlusive ischaemia or vasospastic ischaemia) by measuring

CC expression levels of particular genes (AB199202 to AB199912, encoding  
CC the protein sequences in ABB57020 to ABB57374) or by determining the  
CC expression profile of a gene group comprising these genes. The  
CC expression levels or expression profiles produced by these genes are  
CC used as an indicator when screening for ischaemic condition-improving  
CC drugs or therapeutics for ischaemic diseases. AB199913 and AB199914  
CC represent PCR primers for a mouse ischaemic condition related sequence,  
CC which are used in the exemplification of the present invention.  
XX  
SQ Sequence 176 AA:  
  
Alignment Scores:  
Pred. No.: 106 Length: 176  
Score: 8.00 Matches: 8  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 1.68% Indels: 0  
DB: 23 Gaps: 0  
  
US-10-020-441-1 (1-1477) x ABB57252 (1-176)  
  
QY 1328 GGTGACAGTGTGAGCCTCTCCTC 1351  
ID 130 GlyAspSerGlyGlyProLeuLeu 137  
Db 130 GlyAspSerGlyGlyProLeuLeu 137  
  
RESULT 30  
ABG66758  
ID ABG66758 standard; Protein: 207 AA.  
XX  
AC ABG66758;  
XX  
XX 30-AUG-2002 (first entry)  
XX  
DE Human novel polypeptide #93.  
XX  
XX  
KM Human; inflammatory condition; shock; sepsis; immune response;  
KM cancer; wound healing; central nervous system disease; haematopoiesis;  
KM peripheral nervous system disease; amyotrophic lateral sclerosis; tendon;  
KM myeloid cell disorder; lymphoid cell disorder; platelet disorder; bone;  
KM cartilage; ligament; nerve tissue; ulcer; osteoporosis; osteoarthritis;  
KM bone degenerative disorder; periodontal disease; reperfusion injury;  
KM lung fibrosis; liver fibrosis; autoimmune disorder; bacterial infection;  
KM allergic condition; thrombolysis; thrombosis; coagulation disorder;  
KM fungal infection.  
XX  
XX Homo sapiens.  
XX  
OS  
XX  
PN WO200244340-A2.  
XX  
PD 06-JUN-2002.  
XX  
PF 30-NOV-2001; 2001WO-US47004.  
XX  
PR 30-NOV-2000; 2000US-0028952.  
XX  
PA (HYSE-) HYSEO INC.  
XX  
PI Tang YT, Goodrich RW, Liu C, Zhou P, Asundi V, Wang J, Wang D;  
PI Yamazaki V, Ujwal ML, Drmanac RT;  
XX  
XX WPI; 2002-508509/54.  
XX N-PSDB; ABR94982.  
XX  
PT Novel nucleic acids and polypeptides for diagnosis, treatment of  
PT inflammatory, autoimmune, nervous system, myeloid or lymphoid cell  
PT disorders, cancer and promoting wound healing -  
XX  
XX Claim 10; Page 672; 672pp; English.  
XX  
XX The invention relates to human novel polynucleotides and associated  
CC polypeptides. The polynucleotides and polypeptides are useful for  
CC treating inflammatory conditions such as arthritis, nephritis, Crohn's  
CC disease, ischaemia-reperfusion injury, shock, sepsis, immune responses

CC and cancer and for promoting wound healing. The sequences are used to  
 CC induce the proliferation of neural cells and regeneration of nerve and  
 CC brain tissue, and are useful for the treatment of central and peripheral  
 CC nervous system diseases and neuropathies, such as Alzheimer's disease,  
 CC Parkinson's disease, Huntington's disease and amyotrophic lateral  
 CC sclerosis. The sequences are involved in chemotactic or chemokinetic  
 CC activity, regulation of haematopoiesis, treatment of myeloid or lymphoid  
 CC cell disorders and platelet disorders such as thrombocytopenia,  
 CC regeneration of bone, cartilage, tendon, ligament and/or nerve tissue  
 CC growth, tissue repair, healing of burns, incisions, ulcers, treatment of  
 CC osteoporosis, osteoarthritis, bone degenerative disorders and periodontal  
 CC disease. The sequences of the invention are also useful for gut  
 CC protection or regeneration and treatment of lung or liver fibrosis,  
 CC reperfusion injury in various tissues, immune deficiencies and disorders  
 CC including severe combined immunodeficiency (SCID), bacterial or fungal  
 CC infections, autoimmune disorders e.g. multiple sclerosis and myasthenia  
 CC gravis, allergic conditions such as asthma, thrombolytic or thrombotic  
 CC and coagulation disorders. Sequences ABC65666-ABC66758 represent human  
 CC novel polypeptides of the invention.

XX Sequence 207 AA;

SO Alignment Scores:

Pred. No.: 104 Length: 207  
 Score: 8.00 Matches: 8  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 1.68% Indels: 0  
 DB: 23 Gaps: 0

US-10-020-441-1 (1-1477) x ABC66758 (1-207)

OY 235 TGTGCGTCGCTTGTGCTGCTA 258

DB 149 CysAlaCysValCysValCysVal 156

RESULT 31

AAU54490 ID AAU54490 standard; Protein: 213 AA.

AC AAU54490;

XX 27-FEB-2002 (first entry)

DE Propionibacterium acnes immunogenic protein #15386.

XX SAPHO syndrome; synovitis; acne; pustulosis; hypertosis; osteomyelitis;

KW uveitis; endophthalmitis; bone; joint; central nervous system; ELISA;

KW inflammatory lesion; acne vulgaris; enzyme linked immunosorbent assay;

XX dermatological; osteopathic; neuroprotectant.

OS Propionibacterium acnes.

PN WO200181581-A2.

XX 01-NOV-2001.

PD 20-APR-2001; 2001WO-US12865.

XX 21-APR-2000; 2000US-199047P.

PR 02-JUN-2000; 2000US-208841P.

PR 07-JUL-2000; 2000US-216747P.

XX (CORI-) CORIXA CORP.

DR WPI: 2001-616774/71.

DR N-PSDB: AAS59564.

PT Propionibacterium acnes polypeptides and nucleic acids useful for  
 PT vaccinating against and diagnosing infections, especially useful for

PT treating acne vulgaris -

XX Example 1: SEQ ID No 15685; 1069pp; English.

XX Sequences AAU39105-AAU68017 represent Propionibacterium acnes immunogenic  
 CC polypeptides. The proteins and their associated DNA sequences are used in  
 CC the treatment, prevention and diagnosis of medical conditions caused by  
 CC P. acnes. The disorders include SAPHO syndrome (synovitis, acne,  
 CC pustulosis, hypertosis and osteomyelitis), uveitis and endophthalmitis.  
 CC P. acnes is also involved in infections of bone, joints and the central  
 CC nervous system, however it is particularly involved in the inflammatory  
 CC lesions associated with acne vulgaris. A method for detecting the  
 CC presence or absence of P. acnes in a patient comprises contacting a  
 CC sample with a binding agent that binds to the proteins of the invention  
 CC and determining the amount of bound protein in the sample. The  
 CC polypeptides may be used as antigens in the production of antibodies  
 CC specific for P. acnes proteins. These antibodies can be used to  
 CC downregulate expression and activity of P. acnes polypeptides and  
 CC therefore treat P. acnes infections. The antibodies may also be used as  
 CC diagnostic agents for determining P. acnes presence, for example, by  
 CC enzyme linked immunosorbent assay (ELISA).  
 CC Note: The sequence data for this patent did not form part of the printed  
 CC specification, but was obtained in electronic format directly from WIPO  
 CC at ftp.wipo.int/pub/published\_pct\_sequences.

XX Sequence 213 AA;

SO Alignment Scores:  
 Pred. No.: 104 Length: 213  
 Score: 8.00 Matches: 8  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 1.69% Indels: 0  
 DB: 22 Gaps: 0

US-10-020-441-1 (1-1477) x AAU54490 (1-213)

OY 858 CTCACATCAGTCACCTCAGCTCA 835

DB 67 LeuThrSerArgHisLeuThrSer 74

RESULT 32

AAU53058 ID AAU53058 standard; peptide: 222 AA.

AC AAU53058;

XX 09-OCT-1994 (first entry)

DE Human mature cathepsin-G peptide.

XX Cathepsin-G; chymotrypsin-like cationic protein; gingivitis;

KW periodontitis; antibiotic; antimicrobial; cytostatic.

XX Synthetic.

OS Key Location/Qualifiers

XX Peptide 1..5

XX Peptide 1..20

XX Peptide /label= CG 1-20

XX Peptide /note= "claim 7, page 40"

XX Peptide /label= CG 21-40

XX Peptide /label= CG 41-60

XX Peptide /label= CG 77-82

XX Peptide /label= CG 77-83

XX Peptide /label= CG 77-96

XX Peptide /label= CG 77-96

XX Peptide /label= CG 77-96

FT	Peptide	/label= CG 97-116
FT		61..80
FT		/label= CG 61-80
FT	Peptide	/note= "claim 1, page 39"
FT		117..129
FT		/label= CG 117-129
FT	Peptide	117..136
FT		/label= CG 117-136
FT	Peptide	/note= "claim 1, page 39"
FT		122..136
FT		/label= CG 122-136
FT	Peptide	/note= "claim 1, page 39"
FT		127..136
FT		/label= CG 127-136
FT	Peptide	/note= "claim 1, page 39"
FT		137..156
FT		/label= CG 137-156
FT	Peptide	157..176
FT		/label= CG 157-176
FT	Peptide	177..197
FT		/label= CG 177-197
FT	Peptide	196..223
FT		/label= CG 196-223
FT	Misc-difference	/note= "claim 1, page 39"
FT		181
FT		/note= "target of phosphorylation by DFP"
FT	Misc-difference	44
FT		/note= "charge relay system amino acid"
FT		88
FT	Misc-difference	/note= "charge relay system amino acid"
FT		180
FT	Misc-difference	/note= "charge relay system amino acid"
FT		174
FT		/note= "AA lining Cat-G primary specificity pocket"
FT		177
FT	Misc-difference	/note= "AA lining Cat-G primary specificity pocket"
FT		178
FT	Misc-difference	/note= "AA lining Cat-G primary specificity pocket"
FT		179
FT	Misc-difference	/note= "AA lining Cat-G primary specificity pocket"
FT		194
FT	Misc-difference	/note= "AA lining Cat-G primary specificity pocket"
FT		195
FT	Misc-difference	/note= "AA lining Cat-G primary specificity pocket"
FT		196
FT	Misc-difference	/note= "AA lining Cat-G primary specificity pocket"
FT		203
FT		/note= "AA lining Cat-G primary specificity pocket"
XX		
PN	W09407523-A.	
PD	14-APR-1994.	
XX		
XX		
PF	01-OCT-1993;	93MO-US09414.
XX		
PR	02-OCT-1992;	92US-0956848.
XX		
PA	(UYEM-) UNIV EMORY.	
PA	(UYGE-) UNIV GEORGIA RES FOUND INC.	
PI	Pohl J, Shafer W;	
XX		
DR	WPI; 1994-135222/16.	
XX		
PT	New peptide(s) based on human cathepsin G - having antimicrobial activity, used partic for treating bacterial infection, esp	
PT	gingivitis or periodontitis	
XX		
PS	Disclosure; Fig.1; 4App; English.	
CC	Peptides derived from this sequence can be used for inhibiting growth or controlling infection by a bacterium, particularly for treating gingivitis and/or periodontitis. The peptides	
CC		

CC	can also be used in sterile solutions, herbicidal solutions,
CC	disinfectants and oil recovery fluids.
XX	
sq	Sequence      222 AA:
Alignment Scores:	
Pred. No.:	104
Score:	8.00
Percent Similarity:	100.00%
Best Local Similarity:	100.00%
Query Match:	1.68%
DB:	15
US-10-020-441-1 (1-1477) x AAR53058 (1-222)	
OY	1328 GGTCAGTGGTGACCTCCTC 1351
Db	178 GIYASPERGIYGLProteinU 185
RESULT 33	
AAW01165	
XX	AAW01165 standard; Protein: 223 AA.
AC	AAW01165;
XX	
DT	12-MAR-1997 (first entry)
XX	
DE	Serine protease PfSP4-223.
XX	
KM	Flea; midgut; serine protease; PfSP4-223; vaccine;
KM	domestic animal; Infestation; insecticide; protease-inhibitor;
KW	controlled release formulation; synergist.
XX	
OS	Siphonaptera sp.
XX	
Key	Location/Qualifiers
RH	1..4
FT	/note= "Signal peptide"
FT	1..64
FT	/note= "Sequence AAW01196 (PfSP17, claim 71)"
FT	4..28
FT	/note= "Similar to part of sequence AAW01187 (claim 71)"
FT	Peptide
FT	5..8
FT	/note= "Conserved serine protease motif"
FT	/note=
FT	5..28
FT	"Sequence AAW01182 (claim 71)"
FT	/note=
FT	5..29
FT	"Sequence AAW01178 (claim 71)"
FT	41..47
FT	/note=
FT	Active-site
FT	46
FT	"Conserved region around active site His"
FT	/note=
FT	Misc-difference
FT	90
FT	"Active site His residue"
FT	/note=
FT	Peptide
FT	124..126
FT	"Conserved Asp residue"
FT	/note=
FT	Peptide
FT	124..175
FT	"Conserved serine protease motif"
FT	/note=
FT	Misc-difference
FT	152
FT	"Sequence AAW01169 (claim 71)"
FT	/note=
FT	Misc-difference
FT	165
FT	"Conserved Cys residue"
FT	/note=
FT	Peptide
FT	174..182
FT	"Conserved Cys residue"
FT	/note=
FT	Active-site
FT	179
FT	"Conserved region around active site Ser"
FT	/note=
FT	Peptide
FT	175..182
FT	"Active site Ser residue"
FT	/note=
XX	"Conserved sequence (AAW01226)" ;
PN	WO9611706-A1.
PD	25-APR-1996.
XX	

PF	18-OCT-1995;	95MO-US14442.
XX		
PR	07-JUN-1995;	95US-0485455.
PR	18-OCT-1994;	94US-0326773.
PR	07-JUN-1995;	95US-0482130.
PR	07-JUN-1995;	95US-0484211.
PR	07-JUN-1995;	95US-0485443.
XX		
PA	(PARA-) PARAVAX INC.	
PA	(HESK-) HESKA CORP.	
XX		
PI	Arfsten A, Dale B, Frank GR, Grievie RB, Heath A;	
PI	Hunter SW, Rushlow KE, Stiegler GL, Yamanaka M;	
XX		
DR	WPI: 1996-221762/22.	
DR	N-PSDB: AAT40818.	
XX		
P7	DNA encoding flea serine protease and aminopeptidase - useful in	
P7	vaccines to protect animals from flea infestation.	
XX		
PS	Claim 71: Page 137; 241pp: English.	
XX		
CC	This sequence (PfSP4-223) represents a flea midgut serine protease	
CC	encoded by the nfsP4-672 gene, isolated from a flea cDNA library by	
CC	homology with conserved serine protease sequences (e.g. AA001226).	
CC	Although the entire sequence is not highly conserved to that of	
CC	known serine proteases, it contains a conserved IVGQ sequence, and	
CC	active site His and Ser regions, a conserved GMG sequence, and	
CC	conserved Asp and Cys residues. The sequence contains sequence	
CC	AA001169 (PfSP4-52), which spans 2 conserved serine protease	
CC	sequences, N-terminal sequences AA001196 (PfSPI7), AA001178 and	
CC	AA001182, and a sequence similar to part of AA001187. The protein shows	
CC	homology to an Anopheles gambiae chymotrypsin-I precursor. The	
CC	protein may be used in a vaccine for protection of domestic animals	
CC	from flea infestation, or in isolation of protease-inhibitors, which	
CC	may be used in controlled release formulations to reduce the flea	
CC	burden on and around the animal. The inhibitors may be included in	
CC	insecticidal compositions to increase efficacy of other active	
CC	agents, by reducing proteolytic activity in the flea midgut.	
XX		
SQ	Sequence 223 AA:	
	Alignment Scores:	
	Pred. NO.: 104	Length: 223
	Score: 8.00	Matches: 8
	Percent Similarity: 100.00%	Conservative: 0
	Best local Similarity: 100.00%	Mismatches: 0
	Query Match: 1.66%	Indels: 0
	Db: 17	Gaps: 0
US-10-020-441-1 (1-1477) x AA001165 (1-223)		
OY	1328 GGTGCAGTGGTGACCTTCCTC 1351	
Db	177 GLYASPSERGLGYLPROLEULEU 184	
RESULT 34		
ID	ABB65927 standard; Protein: 223 AA.	
AB	ABB65927	
AC	ABB65927:	
XX		
DT	26-MAR-2002 (first entry)	
DE	Drosophila melanogaster polypeptide SEQ ID NO 24573.	
XX		
KW	Drosophila: developmental biology; cell signalling; insecticide;	
pharmaceutical.		
XX		
OS	Drosophila melanogaster.	
XX		
PN	WO200171042-A2.	
XX		



```

XX  WPI: 1996-484995/48.
DR  N-PSDB: AAT47974.
XX
XX  New recombinant production of active chymase - using a fusion
PT  protein of a non-chymase protein attached to chymase which is
PT  refolded; for producing antibodies and screening for inhibitors
XX
PS  Claim 1; Column 31-34; 31pp; English.
XX
XX  This is the amino acid sequence of the catalytic domain of the mutant
CC  human chymase designated clone 3. The sequences encoding clones 3, 7
CC  (AAT47975) and the wild type sequence (AAT47973) were used to generate
CC  chymase fusion proteins for easier purification of chymase. The use of
CC  a chymase fusion is because chymase produced by itself is
CC  proteolytically active so tends to be unstable in the cell. The fusion
CC  portion of the protein renders the chymase inactive until it is removed
CC  by cleavage. The clones 3 and 7 differ from the wild type in having an A
CC  to G transition mutation resulting in a Lys to Glu residue change at pos.
CC  96, a T to C transition resulting in a Ser to Pro change at pos. 50
CC  (clone 3) and an A to G transition resulting in a Lys to Glu at pos. 204
CC  (clone 7). The purified chymase can be used to generate chymase-specific
CC  antibodies and to identify chymase inhibitors.
XX
SQ  Sequence 226 AA;

Alignment Scores:
Pred. No.: 104      Length: 226
Score: 8.00        Matches: 8
Percent Similarity: 100.00%  Conservative: 0
Best Local Similarity: 100.00%  Mismatches: 0
Query Match: 1.68%  Indels: 0
DB: 17            Gaps: 0

US-10-020-441-1 (1-1477) x AAW09319 (1-226)

QY  1328 GGTGACAGTGTGACCTCTCCTC 1351
    |||||||
DB  180 GLYASPSERGLYGLYProleuleu 187

RESULT 38
AAW09320
ID  AAW09320 standard; Protein; 226 AA.
XX
AC  AAW09320;
XX
DT  14-MAR-1997 (first entry)
XX
DE  Human chymase clone 7 catalytic domain.
XX
KW  Wild type; human; chymase; fusion protein; proteolytically; unstable;
KW  inactive; cleavage; transition mutation; antibody; inhibitor.
XX
OS  Homo sapiens.
XX
PN  US5567602-A.
XX
PD  22-OCT-1996.
XX
PF  12-AUG-1992; 92US-0929198.
XX
PR  12-AUG-1992; 92US-0929198.
XX
PA  (ARRI-) ARRIS PHARM CORP.
XX
PI  Clark JM, Shoemaker KR, Warne RL;
XX
DR  WPI: 1996-484995/48.
XX
DR  N-PSDB: AAT47975.
XX
PT  New recombinant production of active chymase - using a fusion
PT  protein of a non-chymase protein attached to chymase which is
PT  refolded; for producing antibodies and screening for inhibitors
XX

```

```

XX  Claim 1; Column 35-38; 31pp; English.
PS
XX
XX  This is the amino acid sequence of the catalytic domain of the mutant
CC  human chymase designated clone 3. The sequences encoding clones 7, 3
CC  (AAT47974) and the wild type sequence (AAT47973) were used to generate
CC  chymase fusion proteins for easier purification of chymase. The use of
CC  a chymase fusion is because chymase produced by itself is
CC  proteolytically active so tends to be unstable in the cell. The fusion
CC  portion of the protein renders the chymase inactive until it is removed
CC  by cleavage. The clones 3 and 7 differ from the wild type in having an A
CC  to G transition mutation resulting in a Lys to Glu residue change at pos.
CC  96, a T to C transition resulting in a Ser to Pro change at pos. 50
CC  (clone 3) and an A to G transition resulting in a Lys to Glu at pos. 204
CC  (clone 7). The purified chymase can be used to generate chymase-specific
CC  antibodies and to identify chymase inhibitors.
XX
SQ  Sequence 226 AA;

Alignment Scores:
Pred. No.: 104      Length: 226
Score: 8.00        Matches: 8
Percent Similarity: 100.00%  Conservative: 0
Best Local Similarity: 100.00%  Mismatches: 0
Query Match: 1.68%  Indels: 0
DB: 17            Gaps: 0

US-10-020-441-1 (1-1477) x AAW09320 (1-226)

QY  1328 GGTGACAGTGTGACCTCTCCTC 1351
    |||||||
DB  180 GLYASPSERGLYGLYProleuleu 187

RESULT 39
AAW64263
ID  AAW64263 standard; Protein; 247 AA.
XX
AC  AAW64263;
XX
DT  24-NOV-1998 (first entry)
XX
DE  Rat mast cell protease II.
XX
KW  MCP-7; mast cell protease 7; tryptase-7; serine protease;
KW  mast cell protease II; rat blood clot; anticoagulant;
KW  myocardial infarction; reocclusion; thromboembolism;
KW  cerebral embolism; thrombosis; therapy.
XX
OS  Rattus sp.
XX
FH  Key      Location/Qualifiers
FT  Binding-site  55..68  /note="putative substrate binding site"
XX
PN  W09824886-A1.
XX
PD  11-JUN-1998.
XX
PF  25-NOV-1997; 97WO-US21620.
XX
PR  04-DEC-1996; 96US-0032354.
XX
PA  (BGHM ) BRIGHAM & WOMENS HOSPITAL.
XX
PI  Stevens RL;
XX
DR  WPI: 1998-333308/29.
XX
PT  New compositions containing tryptase-7, e.g. mouse mast cell
PT  protease-7 - are used to treat clot formation in e.g. myocardial
PT  infarction, reocclusion following angioplasty or pulmonary
PT  thromboembolism
XX

```

PS Example; Page 79-80; 92pp; English.  
XX  
CC This polypeptide comprises the loop regions in the vicinity of the  
CC putative substrate binding site of rat mast cell protease II. The  
CC crystallographic structure of this region was used as a template  
CC structure to model the structure of the substrate-binding pocket  
CC of mouse mast cell protease 7 (MCP-7, see AAW64233). The invention  
CC relates to MCP-7 and related tryptase-7 proteases that can be used  
CC to prevent or inhibit fibrin clot formation. Such proteases can be  
CC used to treat disorders mediated by undesirable thrombus clot  
CC formation such as myocardial infarction and reocclusion following  
CC angioplasty of blood clots associated with pulmonary  
CC thromboembolism, deep vein thrombosis, cerebral embolism, renal  
CC vein and peripheral arterial thrombosis.  
XX  
SQ Sequence 247 AA;  
  
Alignment Scores:  
Pred. No.: 103 Length: 247  
Score: 8.00 Matches: 8  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 1.68% Indels: 0  
DB: 19 Gaps: 0  
  
US-10-020-441-1 (1-1477) x AAW64263 (1-247)  
  
QY 1328 GGTGACAGTGTGGACCTCTCCTC 1351  
DB 200 GLYASPSERGILYGLYPROLEULeu 207  
|||||  
  
RESULT 40  
AAW73878  
ID AAW73878 standard; Protein; 247 AA.  
XX  
AC AAW73878;  
XX  
DT 31-MAR-1999 (first entry)  
XX  
DE Human chymase protein sequence.  
XX  
KW Chymase; human; secretion expression vector; protein production.  
XX  
OS Homo sapiens.  
XX  
FH Key Location/Qualifiers  
FT Peptide 1..21  
FT Protein /note="signal peptide"  
FT Protein 22..247  
FT Protein /note="mature chymase"  
XX  
PN JP10327870-A.  
XX  
PD 15-DEC-1998.  
XX  
PE 03-JUN-1997; 97JP-0145186.  
XX  
PR 03-JUN-1997; 97JP-0145186.  
XX  
PA (GREG ) GREEN CROSS CORP.  
XX  
XX WPI; 1999-135568/12.  
DR N-PSDB; AAX01231.  
XX  
XX Recombinant human chymase - used to prepare the human chymase  
PT protein  
XX  
PS Claim 2; Page 12-13; 16pp; Japanese.  
XX  
CC This sequence is the human chymase protein. The invention relates to  
CC a method for the preparation of active human chymase by transformation of  
CC a yeast host with a secretion expression vector containing DNA encoding  
CC mature human chymase, and culturing of the cells to give active human

CC chymase. The recombinant DNA can be used to produce the human chymase  
CC protein. Using the method, human chymase can be produced easily and cost  
CC effectively.  
XX  
SQ Sequence 247 AA;  
  
Alignment Scores:  
Pred. No.: 103 Length: 247  
Score: 8.00 Matches: 8  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 1.68% Indels: 0  
DB: 20 Gaps: 0  
  
US-10-020-441-1 (1-1477) x AAW73878 (1-247)  
;  
  
QY 1328 GGTGACAGTGTGGACCTCTCCTC 1351  
DB 201 GLYASPSERGILYGLYPROLEULeu 208  
|||||  
  
Search completed: April 10, 2003, 07:50:58  
Job time : 114 secs



GenCore version 5.1.4.p5\_4578  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - protein search, using frame\_plus\_n2p model

Run on: April 9, 2003, 13:16:43 ; Search time 29.5 Seconds

(without alignments)  
4153.257 Million cell updates/sec

Title: US-10-020-441-1

Perfect score: 2686  
Sequence: 1 gtgtcaacctgtgtacacg.....ctgtacgtccatatttga 1477

Scoring table:

BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 225784

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+n2p model -DEV=xlp  
-Q=/cgn2\_1/USFPO.spool/US10020441/runat\_09042003\_121211\_1318/app.query.fasta\_1.1671  
-DB=swiseprotc 40 -QFMT=fastan -SUFFIX=resp -MINMATCH=0.1 -LOOPEXT=0 -LOOPEXT=0  
-UNITS=bits -START=1 -END=-1 -MATRIX=blotsum62 -TRANS=human40.cdi -LIST=45  
-DOCLLEN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=40 -MODE=LOCAL  
-OUTFMT=plco -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000  
-USER=US10020441@cgn 1.1.34 @runat 09042003\_121211\_1318 -NCPUS=6 -ICPU=3  
-NO\_XLPRY -NO\_MMALP -LARGESQUERY -NEG SCORES=0 -WAIT -LONGLOG -DEV TIMEOUT=120  
-WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7  
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Swiseprotc\_40:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1091.5	40.6	264	1	CERC SCHMA
2	123	4.6	275	1	TRB1_HUMAN
3	120	4.5	235	1	TRYD_HUMAN
4	119	4.4	275	1	TRYA_HUMAN
5	118	4.4	270	1	EL3A_HUMAN
6	117	4.4	270	1	TRYT_MERUN
7	117	4.4	275	1	TRB2_HUMAN
8	115.5	4.3	1161	1	DAN4_YEAST
9	113	4.2	351	1	HRPX_PLALO
10	111.5	4.2	270	1	EL3B_HUMAN
11	111	4.1	276	1	MCT6_MOUSE
12	107.5	4.0	370	1	VE28_HV39
13	107	4.0	246	1	MCTX_MOUSE
14	107	4.0	273	1	MCT7_RAT
15	104.5	3.9	5179	1	MCT4_HUMAN
16	104	3.9	246	1	MCT4_RAT
17	104	3.9	266	1	E11_RAT
18	103.5	3.9	560	1	GAT2_YEAST

C	19	103.5	3.9	651	1	VU47_HSV6U	Q06093 human hepe
C	20	103	3.8	275	1	TRYT_PIG	Q9A2d1 sus scrofa
C	21	103	3.9	1169	1	TRX2_YEAST	P36170 saccharomyc
C	22	102.5	3.8	246	1	MCT9_MOUSE	Q35164 mus musculu
C	23	102.5	3.8	567	1	CH13_CANAL	P40954 candida alb
C	24	102	3.8	244	1	MCT2_MOUSE	P15119 mus musculu
C	25	101.5	3.8	691	1	PEP1_YEAST	P26570 saccharomyc
C	26	101.5	3.8	989	1	PRP3_DICDI	P54637 dictyostell
C	27	100.5	3.8	2440	1	NCR1_HUMAN	Q28153 bos sapien
C	28	100	3.7	266	1	E11_BOVIN	Q28153 bos tauru
C	29	100	3.7	273	1	TRYT_SHEEP	Q9XSM2 ovis arie
C	30	100	3.7	274	1	MCT6_RAT	P05043 rattu norv
C	31	99.5	3.7	321	1	TRYG_HUMAN	Q9NRT2 homo sapien
C	32	99	3.7	273	1	MCT7_MOUSE	Q02844 mus musculu
C	33	99	3.7	290	1	MPN_HUMAN	Q9BQJ3 homo sapien
C	34	99	3.7	3178	1	Y8B5_CAEEL	Q09624 caenorhabdi
C	35	98.5	3.7	247	1	MCT2_RAT	P00770 rattu norv
C	36	98.5	3.7	2248	1	CYAI_DROME	P34870 drosophila
C	37	98	3.7	153	1	YH17_YEAST	P36898 saccharomyc
C	38	98	3.7	213	1	MUC3_HUMAN	Q02505 homo sapien
C	39	98	3.6	275	1	TRYT_CANFA	P15944 canis famli
C	40	97.5	3.6	261	1	CATG_MOUSE	P28293 mus musculu
C	41	97.5	3.7	651	1	VU47_HSV6G	P30005 human hepe
C	42	97	3.6	1119	1	AL53_CANAL	O74623 candida alb
C	43	96.5	3.6	253	1	CAC3_BOVIN	P05805 bos tauru
C	44	96.5	3.6	525	1	YBF4_YEAST	P34219 saccharomyc
C	45	96.5	3.6	779	1	MSS4_YEAST	P38994 saccharomyc

# ALIGNMENTS

## RESULT 1

CERC SCHMA STANDARD: PRT: 264 AA.

AC P12546;  
DT 01-OCT-1989 (Rel. 12, Created)  
DT 01-OCT-1989 (Rel. 12, Last sequence update)  
DT 15-JUN-2002 (Rel. 41, Last annotation update)  
DE Cercarial protease precursor (EC 3.4.21.-) (Cercarial elastase).  
OS Schistosoma mansoni (Blood fluke).  
OC Eukaryota; Metazoa; Platyhelminthes; Trematoda; Digenea; Strigeidida;  
OC Schistosomatidae; Schistosomatidae; Schistosoma.  
ON NCBI\_TaxID=6183;  
RX [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=68330818; Pubmed=3166457;  
RX Newport G.R., McKerrow J.H., Hedstrom R., Pettit M., McGarrigle L.,  
RX Barr P.J., Agabian N.;  
RT "Cloning of the proteinase that facilitates infection by schistosome  
parasites."  
RT J. Biol. Chem. 263:13179-13184 (1988).  
CC - FUNCTION: THIS PROTEASE CLEAVES ELASTIN AND THUS FACILITATES  
CC PENETRATION OF SCHISTOSOME PARASITE LARVAE THROUGH ELASTIN-  
CC RICH TISSUE OF THE HOST.  
CC - ENZYME REGULATION: ACTIVATED BY AN AUTOCATALYTIC MECHANISM.  
CC - SUBCELLULAR LOCATION: ACETABULAR (PENETRATING) GLANDS.  
CC - SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collabora-  
CC tion between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial/  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
CC EMBL: J03946; AAA29864.1; -  
CC PIR: A28942; A28942.  
CC HSSP: P20231; IAAO.  
CC MEROPE: S01.144; -  
CC InterPro: IPR001314; Chymotrypsin.  
CC InterPro: IPR001254; Ser protease\_Try.  
CC Pfam: PF00089; trypsin; I.

DR PRINTS; PR00722; CHYMOTRYPSIN.  
 DR SMART; SM00020; TRYP\_SPE; 1.  
 DR PROSITE; PS00240; TRYPSIN\_DOM; 1.  
 DR PROSITE; PS00134; TRYPSIN\_HIS; 1.  
 DR PROSITE; PS00135; TRYPSIN\_SER; FALSE\_NEG.  
 KM Hydrolyase; Serine protease; Zymogen; Signal.  
 FT SIGNAL 1 19  
 FT PROPEP 20 27 POTENTIAL.  
 FT CHAIN 28 264 CERCARIAL PROTEASE. (BY SIMILARITY).  
 FT ACT\_SITE 68 68 CHARGE RELAY SYSTEM (BY SIMILARITY).  
 FT ACT\_SITE 126 126 CHARGE RELAY SYSTEM (BY SIMILARITY).  
 FT ACT\_SITE 218 218 CHARGE RELAY SYSTEM (BY SIMILARITY).  
 SO SEQUENCE 264 AA; 28545 MW; E2E5129A7C5D5010 CRC64;

## Alignment Scores:

Pred. No.: 1,356-92 Length: 264  
 Score: 1091.50 Matches: 239  
 Percent Similarity: 48.58% Conservative: 0  
 Best Local Similarity: 48.58% Mismatches: 3  
 Query Match: 40.64% Indels: 250  
 DB: 1 Gaps: 2

US-10-020-441-1 (1-1477) x CERC\_SCHEMA (1-264)

QY 1 GTGTCAACCTGGTTGATNAGTAGTGGTGAACCTGTGCAACACCGCATGTAATCCCGTTC 60  
 DB 23 ValSerThrTrpLeuIleArgSerGlyIleProValGlnHisProAlaGluPheProPhe 42  
 QY 61 ATGGATTTCTTAACGACAGAGAGAAACAATGTGTACAGSTTCACTACTGTCACAGAGAGA 120  
 DB 43 IleAlaPheLeuThrThrGluArgThrMetCysThrGlySerIleValSerThrArgAla 62  
 QY 121 GTACTCACAGTGTGTCATGTGTGTTGCTCACCATTTGCCAGTATTCGGTAAGATCGA 180  
 DB 63 ValIleThrAlaGlyHisCysValCysSerProLeuProValIleArg----- 78  
 QY 181 CTGAACAACATTGTGCTCATGTAATCATGATTGATTTCAACAGAGTGTTCGGTGTGCG 240  
 DB 78 ----- 78  
 QY 241 TGCCTTTGTTGTTGCTATGACAGTGGGTTTGCATAGATTTTAATGCTATATCTTGTTC 300  
 DB 78 ----- 78  
 QY 301 TTATTTACGTTTCAATTTCTCACACTGAGGAATGGCAACCAAGGATCCATCACCA 360  
 DB 79 -----ValSerPheLeuThrLeuArgGlnGlyAspGlnGlnIleHisGln 95  
 QY 361 CCGTCTGAGTTAAGGTGGCACACAGATACATGCCCTCTTGTATGTCCGACGACAGAG 420  
 DB 96 ProSerGlyValIleValAlaProGlyTyrMetProSerCysMetSerAlaArgGlnArg 115  
 QY 421 AGACCAATGGCACACACACTCATGTGATTCGATTTGCAATTTGTAATGCTGGCTCAAA 480  
 DB 116 ArgProIleAlaGlnThrLeuSerGlyPheAspIleAlaIleValMetLeuAlaGlnMet 135  
 QY 481 GTCAACTTACAGAGTGCATGACAGATGATGATGATGATGATGATGATGATGATGATGAT 540  
 DB 136 ValIleLeuGlnSerGlyIleArgValIleSerLeuProGlnProSerAspIleProPro 155  
 QY 541 CCTGGAACCTGGTGTGTTTCAATTTGTTGTTATGGAAGGATGATTAACGACCGTGAATCC 600  
 DB 156 ProGlyThrGlyValAlaPheIleValGlyTyrGlyArgAspAspMetProSerProSer 175  
 QY 601 CCTAAGATGGTGGATATTTGAAGAAAGTGAAGTGTGTTGTTGATTAACGACATGACTCAG 660  
 DB 176 ArgGlyAspGlnGlyIleLeuIleuys-Lys----- 184  
 QY 661 TCGTCACTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 720  
 DB 184 ----- 184  
 QY 721 TGTCTACCTGATCCGGTGTGTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTG 780

DB 184 ----- 184  
 QY 781 TGACTTTGTGACAGTTTCACTAGACAGATGATTTCCATCCGTCATGTGTGTGAGAGT 840  
 DB 184 ----- 184  
 QY 841 GAGTGACGTGATGTGAGTGAAGTTGAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGA 900  
 DB 184 ----- 184  
 QY 901 GGATGATTGAGACCACTTGGAGAGAGAAAGACTCATGAATAATCTATGCAACAGATGAA 960  
 DB 184 ----- 184  
 QY 961 GTGTGTGTGATCATGAAGTAGAGGGGTCAATGTGTTGATGTGTTGAGAGAGTGTG 1020  
 DB 184 ----- 184  
 QY 1021 AGATGAGAGTGAATCTTGATTCGTAATTAAGTGAATGATGATGATGATGATGATGAT 1080  
 DB 184 ----- 184  
 QY 1081 TGTGTGGTAGTGTAAAGGTGAGATATTGTGCCAGTTGATATTTTGAATTCAGTTGTG 1140  
 DB 184 ----- 184  
 QY 1141 TGTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTG 1200  
 DB 184 ----- 184  
 QY 1201 GTACTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTG 1260  
 DB 185 -----GlyArgAlaThrIleMetGluCysArg 193  
 QY 1261 ACATCGACCAATGGCAATTCATATGTGTGAAGACAGTCAAGATTTTCCGACATTTAC 1320  
 DB 193 GHisAlaThrAsnGlyAspProIleCysValIleGlnAsnPheGlyGlnLeuP 213  
 QY 1321 CGCTCAGGTGACAGTGGGAGCCTCTCCCATCCCTTCAAGGTCCAGTACTCGGTGT 1380  
 DB 213 CAlaProGlyAspSerGlyGlyProLeuLeuProSerLeuGlnGlyProValLeuGlyVal 233  
 QY 1381 CGTATCACATGTTGTGCACACTGCCAAACCTTCCGATATCATTTGTGATGATCCAGTGT 1440  
 DB 233 ValSerHisGlyValThrLeuProAsnLeuProSerIleIleValGluTyrAlaSerVal 253  
 QY 1441 GGCTAGAAATGTTGATTTTGTACGCTCCCATATT 1474  
 DB 253 lAlaArgMetLeuAspPheValArgSerAsnIle 264  
 RESULT 2  
 ID TRBI\_HUMAN STANDARD; PRT; 275 AA.  
 AC Q15663; Q15663; Q9H2Y4;  
 DT 16-OCT-2001 (Rel. 40, Created)  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Trypsin beta-1 precursor (EC 3.4.21.59) (Trypsinase 1) (Trypsinase I).  
 GN TP5B1.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
 CC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
 OX NCBI\_TaxId=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA MEDLINE=90251647; PubMed=2187193;  
 RX Vanderslice P., Ballinger S.M., Tam E.K., Goldstein S.M., Craik C.S.,  
 RA Caughey G.H.;  
 RT "Human mast cell tryptase: multiple cDNAs and genes reveal a multigene  
 RT serine protease family.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 87:3811-3815(1990).  
 RN [2]

Accession	Source	Length	Matches	Conservative	Mismatches	Indels
RA	SEQUENCE FROM N.A.					
RX	MEDLINE=99121069; PubMed=9930877;					
RA	Pallao M., Pejo M.S., Shayesteh L., Bount J.L., Gaughey G.H.;					
RT	"Characterization of genes encoding known and novel human mast cell					
RT	tryptases on chromosome 16p13.3.";					
RL	J. Biol. Chem. 274:3355-3362(1999).					
RN	[3]					
RP	SEQUENCE OF 54-275 FROM N.A. (ISOFORM 2).					
RA	Wang H.W., McNeil H.P., Thomas P.S., Murphy B.N., Webster M.J.,					
RA	Hettlatchell A., King G., Heywood G.D., Huang C., Stevens R.L.,					
RA	Hunt J.E.;					
RT	"Molecular cloning and characterization of novel human tryptase cDNAs					
RT	and splicing variants.";					
RT	Submitted (NOV-1999) to the EMBL/GenBank/DBJ databases.					
CC	-1- FUNCTION: TRYPTASE IS THE MAJOR NEUTRAL PROTEASE PRESENT IN MAST					
CC	CELLS AND IS SECRETED UPON THE COUPLED ACTIVATION-DEGRADATION					
CC	RESPONSE OF THIS CELL TYPE.					
CC	-1- CATALYTIC ACTIVITY: Preferential cleavage: Arg- - , Lys- - , but					
CC	with more restricted specificity than trypsin.					
CC	-1- SUBUNIT: HOMOTETRAMER (BY SIMILARITY).					
CC	-1- SUBCELLULAR LOCATION: RELEASED FROM THE SECRETORY GRANULES UPON					
CC	MAST CELL ACTIVATION.					
CC	-1- ALTERNATIVE PRODUCTS: 2 isoforms; 1 (shown here) and 2; are					
CC	produced by alternative splicing.					
CC	-1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1. TRYPTASE SUBFAMILY.					
CC						
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration					
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -					
CC	the European Bioinformatics Institute. There are no restrictions on its					
CC	use by non-profit institutions as long as its content is in no way					
CC	modified and this statement is not removed. Usage by and for commercial					
CC	entities requires a license agreement (See <a href="http://www.isb-sdb.ch/announce">http://www.isb-sdb.ch/announce</a>					
CC	or send an email to <a href="mailto:license@isb-sdb.ch">license@isb-sdb.ch</a> ).					
CC	-----					
DR	EMBL; M33494; AAC83172.1; -;					
DR	EMBL; M33494; AAA6778.1; -;					
DR	EMBL; AF099144; AAD17860.1; -;					
DR	EMBL; AF206667; AAG35697.1; -;					
DR	HSSP; P20231; 1A0L.					
DR	MEROPS; S01.242; -;					
DR	Genew; HGNC:12019; TPSBL.					
DR	MIM; 191081; -;					
DR	InterPro; IPR001314; Chymotrypsin.					
DR	InterPro; IPR001254; Ser. protease_Try.					
DR	Pfam; PF00089; trypsin; 1.					
DR	PRINTS; PR00722; CHYMOTRYPSIN.					
DR	SMART; SM00020; Tryp_Spc; 1.					
DR	PROSITE; PS02340; TRYPSIN_DOM; 1.					
DR	PROSITE; PS00134; TRYPSIN_HIS; 1.					
DR	PROSITE; PS00135; TRYPSIN_SER; 1.					
KW	Hydrolase; Serine protease; Signal; Glycoprotein; Zymogen;					
KW	Alternative splicing.					
FT	SIGNAL	1	18	POTENTIAL.		
FT	PROPEP	19	30	ACTIVATION PEPTIDE (BY SIMILARITY).		
FT	CHAIN	31	275	TRYPTASE BETA-1.		
FT	ACT_SITE	74	74	CHARGE RELAY SYSTEM.		
FT	ACT_SITE	121	121	CHARGE RELAY SYSTEM.		
FT	ACT_SITE	224	224	CHARGE RELAY SYSTEM.		
FT	DISULFID	59	75	BY SIMILARITY.		
FT	DISULFID	155	230	BY SIMILARITY.		
FT	DISULFID	188	211	BY SIMILARITY.		
FT	DISULFID	220	248	BY SIMILARITY.		
FT	CARBOHYD	132	132	N-LINKED (GLCNAC. . .) (POTENTIAL).		
FT	CARBOHYD	233	233	N-LINKED (GLCNAC. . .) (POTENTIAL).		
FT	VARSPLIC	79	87	MISSING (IN ISOFORM 2).		
SQ	SEQUENCE	275 AA;	30515 MW;	ADC48FDC51F37112		

DB:	:		1	Gaps:		3
US-10-020-441-1	(1-1477)	x	TRB1_HUMAN	(1-275)		
Oy	91	TGTACAGSTTCACTAAGTCTCAACGACGACTACTCACAAGTGTCATTTGTT-----	144			
Db	59	CysglYglsrLleuIlLeHsProGlntPrVAlleuHrhLaalAhIsQvValGlYPro	78			
Oy	145	---TGCTCACCATTCGCCAGTGCATTCCGGGTAAAGAGATGCATAACACATTTGGTCAA	207			
Db	79	AspVallyAspRleuAlaAlaLeuArgValGlInleuArgGluglnHisLeuYTYtgln	98			
Oy	202	TGTATATGCATGATTTCACACAGCAGTGTTCGCGTCGCGTTCGCTTGTGTGTATGC	261			
Db	99	AspIcn-----	100			
Oy	262	AGTGGGTTTGATGATTTTAATGCTATATACTGGCTTAATTCAGTTTCATTTCTC	322			
Db	101	-----LleuLeuProValSerXrgylelleValnHsProGlnPheTyR	111			
Oy	322	ACACTGAGGAATGCGCAGCAACAAGCATCCATCCACCAACGTCGTGAGTTAAAGTGGCA	381			
Db	115	Thrlagln-----	116			
Oy	382	CCAGATATCATGCCCCCTCTGTATGTCGGCAGCAGACAGAGACCATTCGACACACTC	444			
Db	117	-----	117			
Oy	442	AGTGAATTCGATTTGATGATGTATGTCGGCTCAATTAAGTCACTTACAGAGTGAATC	501			
Db	118	IleoglYlaApRIleAlaLeuIleGluIleuGluIuProValnAvlaserHisVal	133			
Oy	502	AGAAGTATCACTCTGCCACACGCCATCGGATATCCCAGCCACTGGAACTGGTGTTCAT	561			
Db	138	HlsThrValThrIeuProProAlaSerGluThrPheProProGluMetProCystrPrVal	157			
Oy	562	GTTGGTTATGGAGAGGATGATTAACAC	588			
Db	158	ThrlgltYrpgIlyAspValAspAsnasp	166			
<b>RESULT 3</b>						
TRYD_HUMAN		STANDARD;	PRT;	235 AA.		
ID	Q9BZJ3; Q9H2Y6; O9S824;					
AC	16-OCT-2001 (Rel. 40, Created)					
DT	16-OCT-2001 (Rel. 40, Last sequence update)					
DT	15-JUN-2002 (Rel. 41, Last annotation update)					
DE	Putative triptase delta precursor (EC 3.4.21.-) (Mast cell m MCP-7-like) (hm MCP-3-like tryptase III).					
GN	TPSD1.					
OS	Homo sapiens (Human).					
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;					
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.					
OX	NCBI_TaxId=9606;					
RN	[1]					
RX	SEQUENCE FROM N.A.					
RX	Pallero M., Fejzo M.S., Shayesteh L., Blount J.L., Caughay G.H.;					
RA	"Characterization of genes encoding known and novel human mast cell					
RT	tryptases on chromosome 16p13.3.";					
RL	J. Biol. Chem. 274:3355-3362(1999).					
RN	[2]					
RX	SEQUENCE FROM N.A.					
RX	Medline=21101554; PubMed=11174199;					
RA	Min H.K., Kambe N., Schwartz L.B.;					
RT	"Human mouse mast cell protease 7-like tryptase genes are					
RL	pseudogenes.";					
RN	J. Allergy Clin. Immunol. 107:315-321(2001).					
RP	[3]					
RA	Wang H.W., McNeil H.P., Thomas P.S., Murphy B.N., Webster M.J.,					
RA	Hettlerbach A., King G., Heywood G.J., Huang C., Stevens R.L.,					
RA	Hunt J.E.;					

```

RT "Molecular cloning and characterization of novel human trypsin cDNAs
RL Submitted (NOV-1999) to the EMBL/GenBank/DDBJ databases.
CC -1- FUNCTION: This seems to be the product of a pseudogene, it is
CC probably not expressed.
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1. TRYPTASE SUBFAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation --
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@sib.ch).
CC -----
DR EMBL, AF098327; AADI7845.1; ALT_INIT.
DR EMBL, AF318074; AAK12909.1; -.
DR EMBL, AF206664; AAG35694.1; -.
DR HSSP, P20231; IANO.
DR MEROPS, S01.054; -.
DR Genew, HGNC:14118; TPSD1.
DR InterPro, IPR001314; Chymotrypsin.
DR InterPro, IPR001254; Ser_protease_Try.
DR Pfam, PF00089; trypsin, 1.
DR PRINTS, PR00722; CHYMOTRYPSIN.
DR SMART, SMO0020; Tryp_Spc, 1.
DR PROSITE, PS00240; TRYPSIN_DOM, 1.
DR PROSITE, PS00134; TRYPSIN_HIS, 1.
DR PROSITE, PS00135; TRYPSIN_SER, 1.
DR Hypothetical protein; Hydrolase; Serine protease; Signal;
KW Zymogen.
FT SIGNAL 1 18 POTENTIAL.
FT PROPEP 19 30 ACTIVATION PEPTIDE (BY SIMILARITY).
FT CHAIN 31 235 PUTATIVE TRYPTASE DELTA.
FT ACT_SITE 74 74 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE 121 121 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE 224 224 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT DISULFID 59 75 BY SIMILARITY.
FT DISULFID 155 230 BY SIMILARITY.
FT DISULFID 188 211 BY SIMILARITY.
FT CARBOHYD 132 132 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 235 AA; 25816 MW; 81FFB6C2A8006B22 CRC64;

Alignment Scores:
Pred. No.: 0.00419 Length: 235
Score: 120.00 Matches: 37
Percent Similarity: 36.09% Conservative: 24
Best Local Similarity: 21.89% Mismatches: 44
Query Match: 4.47% Indels: 64
DB: 1 Gaps: 3

US-10-020-441-1 (1-1477) x TRYD_HUMAN (1-235)

QY 91 TGTACAGSTTCACTAGTCTCAACGAGAGAGACTACAGCTGGTCATTGTGTT----- 144
||| ||||| : : : ||||| |||||
Db 59 CysGlyGlySerLeuIleHisProGlnTrpValIleThrAlaIleHisCysValGluPro 78
145 ----TGTCAACACTGGCAGTGTGGCGGTGAAGAGATCGACTGAACAACTGTGGCTCA 201
||| ||||| : : : ||||| : : : |||||
Db 79 AspIleLeuAspSerLeuAlaIleLeuArgValGlnLeuArgGlnIleHisLeuValTyrGln 98
202 TGTATCATGATTGATTTCAACAGACAGTGTGGCGTGGCGGCTTGTGTGTGTATGTC 261
Db 99 AspGln----- 100

QY 262 AGTGGGTTTGACATGATTTTAAATGTCTATATATCTGATTTTACAGTTTCATTTCTC 321
||| : : : : : |||||
Db 101 -----LeuIleuProValSerArgIleIleValHisProGlnIlePheTyr 114
322 ACACTGAGGAATGGCCACCAACAAAGCATTCATCACCAACCGTCTGTGAGATTAAAGTGGCA 381
: : : : :
Db 115 IleIleGln----- 117

```

```

Cc 382 CCAGGATACATGCGCCCTTGATGTGCGACGACAGAGAGAGCCAAATGCGACAGACTTC 444
Cc 117 ----- 117
Cc 442 AGTCGATTCGATTTGCGATTGTAATGCTGCTCAATAGTGTAACCTTACAGAGTGAATC 501
Cc 118 ThrlGlyAlaSpilleAlaLeuGluGluGluGluGluGluGluGluGluGluGluGluGlu 137
Cc 502 AGAGTATCAGTCGCGACGACGATTCGCCGACCTGAGCACTGGAACGCGTGTTCATT 561
Cc 138 HisThrValHisLeuProPheProAlaSerGluThrPheProProGlyMetProCysTrpVal 157
Cc 562 GTTGCTATGTAAGGATGATTAACGAC 588
Cc 158 ThrGlyTrpGlyAspValAlaPheAsn 166
Cc
Cc RESULT 4
Cc TRYA HUMAN STANDARD; PRT; 275 AA.
Cc ID TRYA_HUMAN STANDARD; PRT; 275 AA.
Cc AC P15157; Q9U011; Q9H2Y5;
Cc DT 01-APR-1990 (Rel. 14, Created)
Cc DT 16-OCT-2001 (Rel. 40, Last sequence update)
Cc DT 16-OCT-2001 (Rel. 40, Last annotation update)
Cc DE Alpha-tryptase precursor (EC 3.4.21.59) (Tryptase 1).
Cc GN TBS1.
Cc OS Homo sapiens [human].
Cc OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Cc OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
Cc OX NCBI_TaxID=9606;
Cc [1]
Cc RP SEQUENCE FROM N.A. (ISOFORM 1).
Cc RC TISSUE=Lung;
Cc RX MEDLINE=90009311; PubMed=2677049;
Cc RA Miller J.S., Westin E.H., Schwartz L.B.;
Cc RT "Cloning and characterization of complementary DNA for human
Cc tryptase.";
Cc RL J. Clin. Invest. 84:1188-1195(1989).
Cc [2]
Cc RP REVISIONS TO 89-93 AND 108.
Cc RA Schwartz L.B.;
Cc RL Submitted (MAR-1990) to the EMBL/GenBank/DBJ databases.
Cc [3]
Cc RP SEQUENCE FROM N.A. (ISOFORM 1).
Cc RX MEDLINE=99121069; PubMed=9920877;
Cc RA Palladino M., Fejzo M.S., Shayesteh L., Blount J.L., Caughey G.H.;
Cc RT "Characterization of genes encoding known and novel human mast cell
Cc tryptases on chromosome 16p13.3.";
Cc RL J. Biol. Chem. 274:3355-3362(1999).
Cc [4]
Cc RP SEQUENCE FROM N.A. (ISOFORMS 1 AND 2).
Cc RC TISSUE=Lung;
Cc RA Wang H.W., McNeil H.P., Thomas P.S., Murphy B.N., Webster M.J.,
Cc RA Hectiarachi A., King G., Heywood G.J., Huang C., Stevens R.L.,
Cc RA Hunt J.E.;
Cc RT "Molecular cloning and characterization of novel human tryptase cDNAs
Cc and splicing variants.";
Cc RL Submitted (NOV-1999) to the EMBL/GenBank/DBJ databases.
Cc [5]
Cc RP SEQUENCE OF 31-50, AND PITUITARY, SEQUENCE OF 31-38.
Cc RC TISSUE=Lung;
Cc RX MEDLINE=87109258; PubMed=3543004;
Cc RA Cromlish J.A., Seidah N.G., Marcinkiewicz M., Hamelin J., Johnson D.A.,
Cc RA Chretien M.;
Cc RT "Human pituitary tryptase: molecular forms, NH2-terminal sequence,
Cc immunocytochemical localization, and specificity with pronormone and
Cc fluorogenic substrates.";
Cc RL J. Biol. Chem. 262:1363-1373(1987).
Cc CC -I- FUNCTION: TRYPTASE IS THE MAJOR NEUTRAL PROTEASE PRESENT IN MAST
Cc CELLS AND IS SECRETED UPON THE COUPLED ACTIVATION-DEGRADULATION
Cc RESPONSE OF THIS CELL TYPE.
Cc CC -I- CATALYTIC ACTIVITY: Preferential cleavage: Arg-|-, Lys-|-, but
Cc CC with more restricted specificity than trypsin.
Cc CC -I- SUBUNIT: HOMOTETRAMER.

```

CC -1- SUBCELLULAR LOCATION: RELEASED FROM THE SECRETORY GRANULES UPON  
CC MAST CELL ACTIVATION.  
CC -1- ALTERNATIVE PRODUCTS: 2 isoforms; 1 (shown here) and 2; are  
CC produced by alternative splicing.  
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1. TRYPTASE SUBFAMILY.  
-----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
-----  
DR EMBL; M30038; AAA6934.1; -;  
DR EMBL; AF098328; AAD17846.1; -;  
DR EMBL; AF206665; AAG35695.1; -;  
DR EMBL; AF206666; AAG35696.1; -;  
DR PIR; A45754; A45754.  
DR HSSP; P20231; IAOI.  
DR MEROPS; S01.015; -;  
DR MEROPS; S01.143; -;  
DR Genew; HGNC:12018; TP51.  
DR MIM; 191080; -;  
DR InterPro; IPR001314; Chymotrypsin.  
DR InterPro; IPR001254; Ser\_protease\_Try.  
DR Pfam; PF00089; trypsin; 1.  
DR PRINTS; PR00722; CHYMOTRYPSIN.  
DR SMART; SM00020; TRYP\_SEC; 1.  
DR PROSITE; PS0240; TRYP\_SIN\_DOM; 1.  
DR PROSITE; PS00134; TRYP\_SIN\_HIS; 1.  
DR PROSITE; PS00135; TRYP\_SIN\_SER; 1.  
KW Hydrolyase; Serine protease; Signal; Glycoprotein; Zymogen;  
KW Polymorphism; Alternative splicing;  
FT SIGNAL 1 18 POTENTIAL.  
FT PROPEP 19 30 ACTIVATION PEPTIDE.  
FT CHAIN 31 275 ALPHA-TRYPTASE.  
FT ACT\_SITE 74 74 CHARGE RELAY SYSTEM (BY SIMILARITY).  
FT ACT\_SITE 121 121 CHARGE RELAY SYSTEM (BY SIMILARITY).  
FT ACT\_SITE 224 224 CHARGE RELAY SYSTEM (BY SIMILARITY).  
FT DISULFID 59 75 BY SIMILARITY.  
FT DISULFID 155 230 BY SIMILARITY.  
FT DISULFID 188 211 BY SIMILARITY.  
FT DISULFID 220 248 BY SIMILARITY.  
FT CARBOHYD 132 132 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 233 233 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT VASAPLIC 79 87 MISSING (IN ISOFORM 2).  
FT VARIANT 15 15 R->P (IN ALPHA-II).  
FT VARIANT 15 15 /FTID=VAR 012102.  
FT VARIANT 221 221 K->Q (IN ALPHA-II).  
FT CONFLICT 215 216 /FTID=VAR 012103.  
FT SEQUENCE 275 AA; 30772 MW; B9BAC4BBCB9ICE75 CRC64;  
-----  
Alignment Scores:  
Pred. No.: 0.0054 Length: 275  
Score: 119.00 Matches: 36  
Percent Similarity: 35.50% Conservative: 24  
Best Local Similarity: 21.30% Mismatches: 45  
Query Match: 4.43% Indels: 64  
Gaps: 3  
-----  
US-10-020-441-1 (1-1477) x TRY1\_HUMAN (1-275)  
QY 91 TGTACAGSTTCACTAGTCTCAACAGAGACTACTGCTGTCATTTGTT----- 144  
DB 59 Cysgylgylserleuilehisprogintrpvalleuthralaahiscysleuiglypro 78  
QY 145 ---TGTCTACCATTCGACGTGATGGGTAAGATCGACTGAACACATTTGCTGCA 201  
DB 79 AspvallvAspLeuAlaThrleuIrrValGlnleuAlaGgIuGlnHisleuIrrYrGln 98  
QY 202 TGTATCATGATTTCACAGACAGAGTGTTCGCTGCGTTCGTTGTGTGTATGC 261

DB 99 Aspgln----- 100  
QY 262 AGTGGTTTGCATAGATTATTAATGCTATATCTGTTGTTTCAAGTTTCATTTCTC 321  
DB 101 -----LeuIrrProValSerAaGIIleIleValHisProGlnPheTyr 114  
QY 322 ACACAGAGMATGGCGACCAACAGCATTCATCCAAACCGTCTGAGTTAAGTGCA 381  
DB 115 IleIleGln----- 117  
QY 382 CCAGATACATCCCTCTTGTATGTTCGGCAGACAGAGACCAATGCGACACATTC 441  
DB 117 ----- 117  
QY 442 AGTGATTCGATTCATTCATTCATTCATTCATTCATTCATTCATTCATTCATTC 501  
DB 118 ThrGlyAlaAspIleAlaIleuIleuGluIleuGluIleuGluIleuGluIleuGlu 137  
QY 502 AGAGTGATCAGTCTGCACAGCAGCATTCGATTCATTCATTCATTCATTCATTCATTC 561  
DB 138 HSTHrvalMetLeuProProAlaSerGluThrRheProProGlyMetProCysIrrVal 157  
QY 562 GTTGTATGGAAGGATGATTAACGAC 588  
DB 158 ThrGlyTrrGlyAspValAspAsp 166  
-----  
RESULT 5  
EL3A\_HUMAN STANDARD; PRT; 270 AA.  
AC P09093; Q9BRW4;  
DT 01-MAR-1989 (Rel. 10, Created)  
DT 01-FEB-1996 (Rel. 33, Last sequence update)  
DT 15-JUN-2002 (Rel. 41, Last annotation update)  
DS Elascase IIIA precursor (BC 3.4.21.70) (Protease E).  
GN ELA3.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
OX NCBI\_Taxid=9606;  
RN (1)  
RP SEQUENCE FROM N.A.  
RC TISSUE=Pancreas;  
RX MEDLINE=88087253; PubMed=2826474;  
RA Shimada Y., Takemura K., Yoshida H., Sato Y., Iijima H.,  
RA Matsuki S., Tanaka J., Ikenaga H., Ogawa M.;  
RT "Molecular cloning of complementary DNA encoding one of the human  
RT pancreatic protease E isozymes";  
RL J. Biochem. 104:259-264(1988).  
RN (3)  
RP SEQUENCE FROM N.A. AND PARTIAL SEQUENCE.  
RC TISSUE=Pancreas;  
RX MEDLINE=89034017; PubMed=2460440;  
RA Shiraau Y., Takemura K., Yoshida H., Sato Y., Iijima H.,  
RA Shimada Y., Miyayama T., Ozawa T., Ikeda N., Ishida A., Tamai Y.,  
RA Matsuki S., Tanaka J., Ikenaga H., Ogawa M.;  
RT "Molecular cloning of complementary DNA encoding one of the human  
RT pancreatic protease E isozymes";  
RL J. Biochem. 104:259-264(1988).  
RN (3)  
RP SEQUENCE FROM N.A.  
RC TISSUE=Prostate;  
RA Strauberg R.;  
RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.  
CC -1- FUNCTION: EFFICIENT PROTEASE WITH ALANINE SPECIFICITY BUT ONLY  
CC LITTLE ELASTOLYTIC ACTIVITY.  
CC -1- CATALYTIC ACTIVITY: Preferential cleavage: Ala-I-Xaa. Does not  
CC hydrolyse elastin.  
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1. ELASTASE SUBFAMILY.  
-----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

-----  
DR EMBL, M18700; AAA66350.1; JOINED.  
DR EMBL, M18693; AAA66350.1; JOINED.  
DR EMBL, M18694; AAA66350.1; JOINED.  
DR EMBL, M18695; AAA66350.1; JOINED.  
DR EMBL, M18696; AAA66350.1; JOINED.  
DR EMBL, M18697; AAA66350.1; JOINED.  
DR EMBL, M18698; AAA66350.1; JOINED.  
DR EMBL, M18699; AAA66350.1; JOINED.  
DR EMBL, D00306; BAA00212.1; JOINED.  
DR EMBL, BC005918; AAH05918.1; JOINED.  
DR PIR, A29934; A29934.  
DR HSSP, P05805; 1PON.  
DR MEROPS, S01.154; -.  
DR Genew, HGNC:15944; ELA3A.  
DR InterPro, IPR001314; Chymotrypsin.  
DR InterPro, IPR001254; Ser\_protease\_Try.  
DR Pfam, PF00089; trypsin\_1.  
DR PRINTS, PR00722; CHYMOTRYPSIN.  
DR SMART, SM00020; Tryp\_Spc. 1.  
DR PROSITE, PS50240; TRYPSIN\_DOM; 1.  
DR PROSITE, PS00134; TRYPSIN\_HIS; 1.  
DR PROSITE, PS00135; TRYPSIN\_SER; 1.  
KW Hydroxylase; Serine protease; Zymogen; Signal; Glycoprotein.  
FT SIGNAL 1 15  
FT PROPEP 16 28  
FT CHAIN 29 270  
FT ACT\_SITE 73 73  
FT ACT\_SITE 123 123  
FT ACT\_SITE 217 217  
FT DISULFID 58 74  
FT DISULFID 117 120  
FT DISULFID 157 223  
FT DISULFID 188 204  
FT DISULFID 213 244  
FT CARBOHYD 114 114  
FT CONFLICT 63 63  
FT CONFLICT 106 106  
FT CONFLICT 174 174  
SQ SEQUENCE 270 AA; 29474 MW; 576DDB255A4A118C CRC64;  
  
Alignment Scores:  
Pred. No.: 0.00664  
Score: 118.00  
Percent Similarity: 34.33%  
Best Local Similarity: 21.39%  
Query Match: 4.39%  
DB: 1  
Gaps: 4  
  
US-10-020-441-1 (1-1477) x ELA3A\_HUMAN (1-270)  
QY 25 GGTGAACCTGTGCACACCGCACT----- 48  
DB 19 GlyProFroSerSerHisSerSerArgValIValHISGLYGLUAspAlaValProTyr 38  
QY 49 GAATTCGGCTTATCGCATTTAAACGACAGAGAAACA-----ANGTGT 93  
DB 39 SerTyrProTyrIrpIValSerLeuGlnTyrGluLysSerCylSerPheTyrHisThrCys 58  
QY 94 ACAGSTTCATAGTCTCAACGACGAGAGAGTCACTGGTGTGTTGTTCTCCACCA 153  
DB 59 GlyGlySerLeuIleAlaProAspTyrPValIValThrAlaGlyHisCysLysSerArgAsp 78  
QY 154 TTGCCAGTATTCGGGTGAAGAGATGACATGAACACATTTGTTGCTCAATGATTCATTG 213  
DB 79 Leu----- 79  
QY 214 ATTTCACAGACAGTGTGTTGGGTGGCGTGGCTTTGTTGTGTATGACAGTGGGTTTGA 273  
DB 79 ----- 79

QY 274 TAGATTTAATGCTTAATTAAGTGTCTTAATTTACAGTTTCACTGACAGAAAT 333  
DB 80 -----ThrTyrGlnValValLeuGlyGluTyrAsnLeuAlaValLysGlu 94  
QY 334 GGGCACAACAAGGATCTCATCAACCAACCGTTGGAGTTAAAGTGGACCAAGATACATG 393  
DB 95 GlyProGluGlnValIleProIleAsnSerGluGlnLeuPheValHisProLeuTyrAsn 114  
QY 394 CCCCTTGTATGTCGGCAGACAGAGAGACCAATTCGACACACATGACATGATTCAT 453  
DB 115 ArgSerCysValAlaCys-----GlyAsnAsp 123  
QY 454 ATTGCATTTAATGCTGCTCAATATGCTCACTTACAGAGTGAATGACAGTATCACT 513  
DB 124 IleAlaIleIleLysLeuSerArgSerArgIleuGlnLeuGlyAspAlaValGlnLeuLaser 143  
QY 514 CTGCCACAGCCATCGGATATCCCGCACCCTGGAATCTGTTTTCATTTGTTATGGA 573  
DB 144 LeuProProAlaGlyAspIleLeuProAsnLysThrProCysTyrIleThrGlyTyrGly 163  
QY 574 AGG 576  
DB 164 Arg 164  
  
RESULT 6  
TRYT\_MERUN  
ID TRYT\_MERUN STANDARD; PRT; 270 AA.  
AC P05342;  
DT 01-OCT-1996 (Rel. 34, Created)  
DT 01-OCT-1996 (Rel. 34, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE Mast cell tryptase precursor (EC 3.4.21.59).  
OS Meriones unguiculatus (Mongolian jird) (Mongolian gerbil).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Gerbillinae;  
OC Meriones.  
OX NCBI\_TaxID=10047;  
[1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=MGS/SEA; TISSUE=Intestine;  
RX MEDLINE=95366971; PubMed=7639711;  
RA Murakumo Y., Ide H., Itoh H., Tomlita M., Kobayashi T.,  
RA Maruyama H., Horii Y., Nawa Y.;  
RT "Cloning of the cDNA encoding mast cell tryptase of Mongolian gerbil,  
RT Meriones unguiculatus, and its preferential expression in the  
RT intestinal mucosa." (1995).  
RL Biochem. J. 309:921-926(1995).  
CC -|- CATALYTIC ACTIVITY: Preferential cleavage: Arg-|-, Lys-|-, but  
CC with more restricted specificity than trypsin.  
CC -|- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1. TRYPTASE SUBFAMILY.  
-----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation-  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

-----  
DR EMBL, D31789; BAA06598.1; -.  
DR HSSP, P20231; IAAO.  
DR MEROPS, S01.143; -.  
DR InterPro, IPR001314; Chymotrypsin.  
DR InterPro, IPR001254; Ser\_protease\_Try.  
DR Pfam, PF00089; trypsin\_1.  
DR PRINTS, PR00722; CHYMOTRYPSIN.  
DR SMART, SM00020; Tryp\_Spc. 1.  
DR PROSITE, PS50240; TRYPSIN\_DOM; 1.  
DR PROSITE, PS00134; TRYPSIN\_HIS; 1.  
DR PROSITE, PS00135; TRYPSIN\_SER; 1.  
KW Hydroxylase; Serine protease; Signal; Glycoprotein.  
FT SIGNAL 1 25  
FT PROPEP 26 28  
FT CHAIN 29 270



```

DR EMBL; M37488; AAA51843.1; -
DR EMBL; M33492; AAA36779.1; -
DR EMBL; M33493; AAA36780.1; -
DR EMBL; S55551; AAD13876.1; -
DR EMBL; AF099143; AAD17859.2; -
DR EMBL; AF099145; AAD17857.1; -
DR EMBL; AF099146; AAD17858.1; -
DR PIR; A37193; A37193.
DR PIR; B35863; B35863.
DR PDB; 1A0J; 31-JAN-94.
DR PDB; 1A0L; 23-MAR-99.
DR MEROPS; S01.027; -.
DR MEROPS; S01.242; -.
DR Genew; HGNC:14120; TPSB2.
DR InterPro; IPR001314; Chymotrypsin.
DR InterPro; IPR001254; Ser_protease_Try.
DR Pfam; PF00089; trypsin; 1.
DR PRINTS; PR00722; CHYMOTRYPSIN.
DR SMART; SM00020; Tryp_Spc; 1.
DR PROSITE; PS0240; TRYPsin DOM; 1.
DR PROSITE; PS00134; TRYPsin HIS; 1.
DR PROSITE; PS00135; TRYPsin_SER; 1.
KM Hydrolase; Serine protease; Signal; Glycoprotein; Zymogen;
KM Polymorphism; 3d-structure.
FT SIGNAL 1 18 POTENTIAL.
FT PROPEP 19 30 ACTIVATION PEPTIDE.
FT CHAIN 31 275 TRYPTASE BETA-2.
FT ACT_SITE 74 74 CHARGE RELAY SYSTEM.
FT ACT_SITE 121 121 CHARGE RELAY SYSTEM.
FT ACT_SITE 224 224 CHARGE RELAY SYSTEM.
FT DISULFID 59 75
FT DISULFID 155 230
FT DISULFID 188 211
FT DISULFID 220 248
FT CARBOXD 223 233 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT VARIANT 51 53 HCP -> RDR (IN BETA-II1).
FT CONFLICT 132 132 /FTID=VAR_012104.
SQ SEQUENCE 275 AA; 30529 MW; 2B27396C51F5C7A0 CRC64;
Alignment Scores:
Pred. No.: 0.00825 Length: 275
Score: 117.00 Matches: 37
Percent Similarity: 34.91% Conservative: 22
Best Local Similarity: 21.89% Mismatches: 46
Query Match: 4.36% Indels: 64
DB: 1 Gaps: 3
US-10-020-441-1 (1-1477) x TRB2_HUMAN (1-275)
QY 91 TGTACAGTTCACTAGTCTCAACGAGACGACTACAGCTGTCATTGTT----- 144
DB 59 CysGlyGlySerLeuIleHisProGlnTrpValLeuThrAlaHisCysValGlyPro 78
QY 145 ---TGCTCAACATTGCGACGATTCGGGTAAAGATCGACTGAACACATTGTTGCTCA 201
DB 79 AspValIlyAspLeuAlaIleuArgValGlnLeuArgGlnIleHisLeuTyrTyrGln 98
QY 202 TGTATCGATTGATTTCACAGACAGTGTTCGCTGTCGCTGTCGCTGTTGTGTATGTC 261
DB 99 AspGln-----LeuLeuProValSerArgIleIleValHisProGlnPheTyr 114
QY 262 AGTGGGTTGCATATGATTATTAATGCTATATGCTTATTTCAGGTTTCATTTTC 321
DB 101 -----LeuLeuProValSerArgIleIleValHisProGlnPheTyr 114
QY 322 ACACTGAGAAATGGCGACCAACAGGACATCCATCAACACCGTCTGAGATTAAAGTGGCA 381
DB 115 ThrAlaGln----- 117
QY 382 CCAAGATTAATGCCCTCTTGTATGTGCGACGACAGAGAGACCAATCGACAGACATC 441
DB 117 ----- 117

```

```

QY 442 AGTGATTGCATATTCATTGTAATGCTGCTCAATGCTCAACTTACAGATGGAATC 501
DB 118 IleGlyAlaAspIleAlaLeuLeuGlnIleuGlnIleuProValIleValSerHisVal 137
QY 502 AGAGTATCATCTGCGCCACAGGACATCGATATCCGCCACCTGGAACTGTTTCATT 561
DB 138 HisThrValThrLeuProProAlaSerGlnThrPheProProGlyMetProCysTrpVal 157
QY 562 GTTGTATGGAAGGATGATTAACGAC 588
DB 158 ThrGlyTrpGlyAspValAspAsnAsp 166
RESULT 8
DAN4_YEAST
ID DAN4_YEAST STANDARD; PRT; 1161 AA.
AC P47179.1
DT 01-FEB-1996 (Rel. 33, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Cell wall protein DAN4 precursor.
GN DAN4 OR YJR151C OR J2223.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RA Scarcez T.;
RL Submitted (SEP-1995) to the EMBL/GenBank/DBJ databases.
RN [2]
RP REGULATION.
RX MEDLINE=21113168; PubMed=11160904;
RA Cohen B.D., Serflin O., Abramova N.E., Davies K.J., Lowry C.V.;
RT "Induction and repression of DAN1 and the family of anaerobic
RT mannoprotein genes in Saccharomyces cerevisiae occurs through a
RT complex array of regulatory sites.";
RL Nucleic Acids Res. 29:799-808(2001).
CC -!- FUNCTION: COMPONENT OF THE CELL WALL (By similarity).
CC -!- SUBCELLULAR LOCATION: Attached to the membrane by a GPI-anchor
CC (potential).
CC -!- PTM: EXTENSIVELY O-GLYCOSYLATED (POTENTIAL).
CC -!- SIMILARITY: BELONGS TO THE SRP1 / TIPI FAMILY.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
DR EMBL; Z49651; CA89684.1; -
DR SGD; S0003912; DAN4.
DR InterPro; IPR000992; SRP1_TIP1.
DR Pfam; PF00660; SRP1_TIP1; 1.
DR PROSITE; PS00724; SRP1_TIP1; 1.
KW Cell wall; Glycoprotein; Membrane; GPI-anchor; Signal.
FT SIGNAL 1 24 POTENTIAL.
FT CHAIN 25 1146 CELL WALL PROTEIN DAN4.
FT PROPEP 1147 1161 REMOVED IN MATURE FORM (POTENTIAL).
FT LIPID 1146 GPI-ANCHOR (POTENTIAL).
SQ SEQUENCE 1161 AA; 118358 MW; 7954C15D9F0CA58 CRC64;
Alignment Scores:
Pred. No.: 0.0165 Length: 1161
Score: 115.50 Matches: 62
Percent Similarity: 37.64% Conservative: 37
Best Local Similarity: 23.57% Mismatches: 127
Query Match: 4.33% Indels: 37
DB: 1 Gaps: 3
US-10-020-441-1 (1-1477) x DAN4_YEAST (1-1161)

```



Qy	1270	GGTGGACATGTCGGATTCATTTATAGTCGGTGGACCTGGTGGTGAACAACAACACCA	121
Dy	116	GIYIEYrThrAlaAlaIeProThrSerThrThrThrThySsrSerThrSer	135
Qy	1210	ACAACAGTACAGTGAACAACAACGTACAGTACAGATGAGAAGAAACAACAACAGGACA	1151
Dy	136	ThrThrPro-ThrThrThrIleThrSerThrThrSerThrThrThrThrProThrTh	155
Qy	1150	AAACAAACACACACAAGTGAATTTGGAAAAATATCACTGGCACAATATCCACCTTTAAC	1091
Dy	155	rSerThrThrSerThrThrProThrThrSerThrThrSerThrThrProThrThrSerTh	175
Qy	1090	TACCACACCAACATATGACACATACATACATGACACTATATATGACAGATCAAGTCA	1031
Dy	175	rThrSerThrThr-----ProThrThr	182
Qy	1030	CTCTCATCTTCACACCTCTCCAAACATATCTGAACACATTTGACCCCTACTGTATGA	971
Dy	182	rSerThrThrSerThrThrProThrThrSerThrThrThrThrThrThrThrThrThr	202
Qy	970	CACACACACTCTTCATGCTTTGATGATATTTTCATGAGTCTTCTCTCCAAAGTGTC	911
Dy	202	hTrhTrSerThr-----ThrProThrThrSerT	211
Qy	910	TCATATCATCCATCATCATCATTCATCCATCCATCCATTAACCTGAACACATCATCAATC	851
Dy	211	hTrhTrSerThrThrProThrThrSerThrThrSerThrThrProThrThrSerThrThrP	231
Qy	850	ACGTACCTCCTACCTCACCAACACATGACCGAGATGGAATCACTCTGCTCACTGAACGT	791
Dy	231	rOTrhTrhTrSerThrThrSerThrThrSerGlnThrSerThrThrLysSerThrThrProThrT	251
Qy	790	CACAAAGTCATCCAAACACAGTGTATTATTCAGAGCTGTGACCAATATCAACACCGAT	731
Dy	251	hTrSerThrThrSerThrThrThrProThrThrSerThrThrProThrThrSerThrThrSerT	271
Qy	730	CAGGTAGACAGACAGACACACAACAAAGATATGCACAGACAGAAATTACTGCATGACATC	671
Dy	271	hTrAlaProThrThrThrSerThrThrSerThrThrSerThrThrThrThrThrThrThrAla	290
Qy	670	TGACTGACGACTGATGATGTCGTTATTATTCACCAACACTCACTTCTTCAATATTCAC	611
Dy	291	ProThrThrSerThrThrThrSerThrThrPheSerThrSerSerAlaSerAla-SerSerVa	310
Qy	610	CATCTTACGAGACGATCAACGGTCGTTATCATCCTTCATCATACCAACAAGAAAAAC	551
Dy	310	IlIleSerThrThrAlaThrThrThrSerThrThrPheAlaSerIleThrThrProAlaThrSe	330
Qy	550	CAGTTCAGGATGGGAGGATTCGATGGATGGCTGGACAGACTGATCACTCTGATTCACCT	491
Dy	330	rThrAla-----SerThrAlaSerThrThrSerSerVa	341
Qy	490	G 490	
Dy	341	1 341	

RESULT 9

HRPX_PLALO	STANDARD;	PRT;	351 AA.
ID_HRPX_PLALO			
AC_P04929;			
DT 13-AUG-1987 (Rel. 05, Created)			
DT 13-AUG-1987 (Rel. 05, Last sequence update)			
DT 15-JUN-1999 (Rel. 38, Last annotation update)			
DE Histidine-rich glycoprotein precursor.			
OS Plasmodium falciparum.			
OC Eukaryote; Alveolata; Apicomplexa; Haemosporida; Plasmodium.			
OX NCBI_TaxId=5853;			
RA			
RP SEQUENCE FROM N.A.			
RP MEDLINE=85061618; PubMed=6095114;			
RA Ravetch J.V., Feder R., Pavlovic A., Blobel G.,			
RT "Primary structure and genomic organization of the histidine-rich			

[illegible]



QY 331 AATGCGACCAAGCATCCATCCATCCACCGCTGAGTTAAGTGCACACGAGTAC 390  
 DB 94 GUGLYPGGLUGLHVALLLEPROLEAENSERGLYASPLEUPHEVALHISPROLEUTRP 113  
 QY 391 ATGCCCTCTTGTATGTCGGCAGACAGAGACCAATGCACACTCAGTGTATTC 450  
 DB 114 AenArgSerCyValAlaCyS-----GlyAen 122  
 QY 451 GATTTGCAATGTATGCTGGCTCAATGTCACTTACAGAGTGATGATGATC 510  
 DB 123 AapllelealeullelleyleuserargserAlaGlnleuGlyASpAlaValGlnleuAla 142  
 QY 511 AGTGTGCCACAGCCATCCGATATCCCGCAGCTGGAATGCTGTTTCATTTGTTAT 570  
 DB 143 SerleupProAlaGlyAaplleleupProaangluThrProCySerIleThrglyTrp 162  
 QY 571 GGAAGG 576  
 DB 163 GlyArg 164

RESULT 11  
 MCT6\_MOUSE  
 ID MCT6\_MOUSE STANDARD; PRT; 276 AA.  
 AC P21845; Q61962;  
 DT 01-MAY-1991 (Rel. 18, Created)  
 DT 01-AUG-1991 (Rel. 19, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE Mast cell protease 6 precursor (EC 3.4.21.55) (MCP-6) (tryptase).  
 GN MCTP6.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OC NCBI\_Taxid=10090;  
 [1]  
 RN SEQUENCE FROM N.A.  
 RX MEDLINE=91139682; PubMed=1995638;  
 RA Reynolds D.S., Gurley D.S., Austen K.F., Serafin W.E.;  
 RT "Cloning of the cDNA and gene of mouse mast cell protease-6.  
 RT Transcription by progenitor mast cells and mast cells of the  
 RT connective tissue subclass.",  
 RL J. Biol. Chem. 266:3847-3853 (1991).  
 RN [2]  
 RN SEQUENCE FROM N.A.  
 RC STRAIN=Leaden X A1;  
 RX MEDLINE=94023807; PubMed=8210998;  
 RA Huang R., Abbrink M., Gohl A.E., Nilsson G., Aveskog M., Larsson L.G.,  
 RA Nilsson K., Hellman L.;  
 RT "Expression of a mast cell tryptase in the human monocytic cell lines  
 RT U-937 and Mono Mac 6.";  
 RL Scand. J. Immunol. 38:359-367 (1993).  
 RN [3]  
 RN SEQUENCE FROM N.A. (SHORT ISOFORM).  
 RC STRAIN=Leaden X A1;  
 RX MEDLINE=95048582; PubMed=7959952;  
 RA Huang R., Hellman L.T.;  
 RT "Genes for mast-cell serine protease and their molecular evolution.";  
 RL Immunogenetics 40:397-414 (1994).  
 RN [4]  
 RN SEQUENCE OF 32-54.  
 RX MEDLINE=90222202; PubMed=2326280;  
 RA Reynolds D.S., Stevens R.L., Lane W.S., Carr M.H., Austen K.F.,  
 RA Serafin W.E.;  
 RT "Different mouse mast cell populations express various combinations  
 RT of at least six distinct mast cell serine proteases.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 87:3230-3234 (1990).  
 CC -1- FUNCTION: TRYPTASE IS THE MAJOR NEUTRAL PROTEASE PRESENT IN MAST  
 CC CELLS, AND IS SECRETED UPON THE COUPLED ACTIVATION-DEGRANULATION  
 CC RESPONSE OF THIS CELL TYPE.  
 CC -1- CATALYTIC ACTIVITY: Preferential cleavage: Arg-|-, Lys-|-, but  
 CC with more restricted specificity than trypsin.  
 CC -1- ALTERNATIVE PRODUCTS: 2 isoforms; a long form (shown here) and a  
 CC short form; are produced by alternative splicing. The short form

CC is probably a non functional variant.  
 CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1. TRYPTASE SUBFAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.ebi.ac.uk/announce/>  
 CC or send an email to [license@ebi.ac.uk](mailto:license@ebi.ac.uk)).

DR EMBL, M57626; AAA39988.1; -;  
 DR EMBL, M57625; AAA39987.1; -;  
 DR EMBL, L31853; AAA39725.1; -;  
 DR EMBL, X78542; CAA5288.1; -;  
 DR PIR, A38654; A38654.  
 DR PIR, D35646; D35646.  
 DR HSRP, P20231; IAAO.  
 DR MEROPS, S01.025; -.  
 DR MCD, MGI:96942; Mcpt6.  
 DR InterPro: IPR001314; Chymotrypsin.  
 DR InterPro: IPR001254; Ser protease\_Try.  
 DR Pfam, PF00089; trypsin\_1.  
 DR PRINTS; PR00722; CHYMOTRYPSIN.  
 DR SMART; SM00020; TRYD\_SPC; 1.  
 DR PROSITE; PS50240; TRYPSIN\_DOM; 1.  
 DR PROSITE; PS00134; TRYPSIN\_HIS; 1.  
 DR PROSITE; PS00135; TRYPSIN\_SER; 1.  
 KM Hydrolase, Serine protease; Signal; Glycoprotein; Zymogen;  
 KM Alternative splicing.  
 FT SIGNAL 1 21  
 FT PROPEP 22 31  
 FT CHAIN 32 276  
 FT ACT\_SITE 75 75  
 FT ACT\_SITE 122 122  
 FT ACT\_SITE 225 225  
 FT DISULFID 60 76  
 FT DISULFID 156 231  
 FT DISULFID 189 212  
 FT DISULFID 221 249  
 FT CARBOHYD 223 133  
 FT VARSPLIC 223 230  
 FT VARSPLIC 231 276  
 SQ SEQUENCE 276 AA; 30927 MW; 525B2C9A04A72200 CRC64;

Alignment Scores:  
 Pred. No.: 0.0295 Length: 276  
 Score: 111.00 Matches: 41  
 Percent Similarity: 34.86% Conservative: 20  
 Best Local Similarity: 23.43% Mismatches: 38  
 Query Match: 4.13% Indels: 76  
 DB: 1 Gaps: 5

US-10-020-441-1 (1-1477) x MCT6\_MOUSE (1-276)

QY 91 TGTAAGSTTCACTGATGTCACAGACAGTCAAGCTGATGTT----- 144  
 DB 60 CySGlyGlySerLeuIleHisProGlnTrpValLeuThrAlaAlaHisCyValGlyPro 79  
 QY 145 -----TGCTCAACATTCGATGATGGGTGAAGATGATGATGAAACATTT----- 192  
 DB 80 HisIleLysSerPro---GlnleupheargValGlnleuAargGlnTrpLeuTrpTyr 98  
 QY 193 -----GTTGCTCAATGTATGATGATTTTCACAGACAGTGTTCGTCGTCGC 243  
 DB 99 GlyAepGlnleuLeuSerleuAenArgIleVal----- 109  
 QY 244 GTTGTGTTTGTGATGATGAGGGTTTCATATGATTTAATGTATATCTGTTGTTTA 303  
 DB 109 ----- 109  
 QY 304 TTTGAGTTTCAATTTCTCACATGAGAGATGCGACCAAGAGCATCATCACCAACCG 363

```

Db 109 ----- 109
QY 364 TCTGAGTTAAGTGACACAGATACATGCCCTCTGTATGTGCGACAGACAGAGA 423
Db 110 -----ValHisProHisTyrThrAla----- 117
QY 424 CCMAATCGACAGACACTGATTCGATTGCAATTGCAATTGATGCTGCTCAATGCTC 483
Db 118 -----GluGlyGlyAlaSerValAlaLeuLeuGluValProVal 132
QY 444 AACTTACAGAGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 543
Db 133 AsnValSerThrHisIleHisProIleSerLeuProIleSerValIleThrPhePro 152
QY 544 GGAATCGGTGTTTCAATGTTGTTATGAGAGGATGATGATGATGATGATGATGATGAT 588
Db 153 GlyThrSerCysTrpValThrGlyTrpGlyAspIleAspAsnAsp 167

RESULT 12
VE2_HPV3          STANDARD;          PRT;          370 AA.
ID   VE2_HPV3
AC   P24830.
DT   01-MAR-1992 (Rel. 21, Created)
DT   01-MAR-1992 (Rel. 21, Last sequence update)
DT   15-JUL-1998 (Rel. 36, Last annotation update)
DE   Regulatory protein E2.
GN   E2.
OS   Human papillomavirus type 39.
OC   Viruses; dsDNA viruses, no RNA stage; Papillomaviridae;
OC   Papillomavirus.
OX   NCBI_TaxID=10588;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=91135017; PubMed=1847266;
RA   Volpers C., Strebeck R.E.;
RT   "Genome organization and nucleotide sequence of human papillomavirus
RT   type 39.";
RL   Virology 181:419-423(1991).
CC   -!- FUNCTION: E2 REGULATES VIRAL TRANSCRIPTION AND DNA REPLICATION.
CC   IT BINDS TO THE E2RE RESPONSE ELEMENT (5'-ACCCNNNNNGGT-3') PRESENT
CC   IN MULTIPLE COPIES IN THE REGULATORY REGION. IT CAN EITHER
CC   ACTIVATE OR REPRESS TRANSCRIPTION DEPENDING OF E2RE'S POSITION
CC   WITH REGARDS TO PROXIMAL PROMOTER ELEMENTS. REPRESSION OCCURS
CC   BY STERICALLY HINDERING THE ASSEMBLY OF THE TRANSCRIPTION
CC   INITIATION COMPLEX. THE E1-E2 COMPLEX BINDS TO THE ORIGIN OF DNA
CC   REPLICATION.
CC   -!- SUBUNIT: BINDS DNA AS A DIMER.
CC   -!- SUBCELLULAR LOCATION: Nuclear.
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@sib-sib.ch).
CC   EMBL: M62849; AAA47053.1; -.
DR   PIR; D38502; W2M139.
DR   HSP; P17383; IDHM.
DR   InterPro; IPR000427; E2_C.
DR   InterPro; IPR001866; E2_N.
DR   Pfam; PF00508; E2_N; 1.
DR   Pfam; PF00511; E2_C; 1.
DR   ProDom; PD000672; E2_C; 1.
DR   ProDom; PD000678; E2_N; 1.
KW   Early protein; Transcription regulation; Activator; DNA-binding;
KW   Trans-acting factor; DNA replication; Repressor; Nuclear protein.
SQ   SEQUENCE 370 AA; 42697 MW; D8EF12CDE1C17F17 CRC64;

```

Alignment Scores: 0.0669 Length: 370  
 Pred. No.: 107.50 Matches: 61  
 Score:

```

Percent Similarity: 36.10% Conservative: 26
Best Local Similarity: 25.31% Mismatches: 75
Query Match: 4.03% Indels: 79
DB: 1 Gaps: 14

US-10-020-441-1 (1-1477) x VE2_HPV3 (1-370)
QY 1384 TACGACACCGAGTACGTGACCTTGAGAGGATGGAGAGAGGATCCACCATGTCACCTGG 1325
Db 92 TyrAsnThrGluGluTrpThrLeuLysAsp-----ThSerAsnGluLeuTrp 107
QY 1334 AGCGGTACATGTCGGAATTCGACCTGCTTCACATATAGATTGCCATT----- 1271
Db 108 His-----ThrGlnProLysGlnCysPheLysGlnGlyThrThrValGluVal 124
QY 1270 -----GGTGGCATGCGGCATTCCATTAT----- 1247
Db 125 TrpTyrAspGlyAspLysCysAsnAlaMetCenTyrValLeuTrpGlyAlaIleTyrTyr 144
QY 1246 --AGTCGCTCGACCCCTGTGTGTCMAACMAACMAACMAACMAACMAACMAACMAAC 1190
Db 145 LysAsnAsnIleAspIleTrpCysLys-----ThrGluGlyCysValAspTyrTrpGly 162
QY 1189 GTACAGTACAGATGAGAAAGAACAAA-----CACACGGAGCA 1151
Db 163 IleTyrTyrMetAsnGlnHisIleuLysValTyrTyrGluValPheIleGlnAspAlaGlu 182
QY 1150 AACAACAAACACACAGTGAATTGCAAAATATCAATCGGACCAATATCCACCTTTAGAC 1091
Db 183 ArgTyrGlyThrSerGlyLysTrpGlu-----ValHis 193
QY 1090 TACCACACACMAATAGTCCACATA--CAATCAGTGTGACCTATATTCAGACATCAG 1034
Db 194 TyrAsnGlyAsn-IleIleHisCysProAspSer-MetCysSerThrSerAspGlySerV 213
QY 1033 TC--ACTTCATCATTCACCACT--CTCCAAACACATCTCAAAACACATTACCCCCCTA 980
Db 213 aLProThrThrGluLeuThrThrGluLeuSerAsnThrThrAlaThrHisSerThrAlaT 233
QY 979 CTTC--TGTACACACACACTTCATCGTTGATAGATATTCATGAGCTTCTCTC 923
Db 233 hTrhProCysThrGlnLysThr-----IleProP 243
QY 922 CTCCAAGT----- 915
Db 243 roProSerArgLysArgProArgGlnCysAlaValThrGluProThrGluProAspGlyV 263
QY 914 -----GGTTCATCATCCCATACATGATTCATCCCATTCATTCATC 866
Db 263 aLSerLeuAspHisIleuAsnAsnProLeuHisSerAsnSerThrGlyHisAsnThrArgA 283
QY 865 AACTCACTCAATCATCACTCACTCACT-----CACCAACACAAATGACGAGATGA 813
Db 283 xGlyrLeuSerCysGlyAsnThrThrProIleIleHisLeuLysGlyAspLysAsnGly 302

RESULT 13
MCTX_MOUSE
ID   MCTX_MOUSE          STANDARD;          PRT;          246 AA.
AC   O00356.
DT   01-JUL-1993 (Rel. 26, Created)
DT   01-JUL-1993 (Rel. 26, Last sequence update)
DT   15-JUL-1998 (Rel. 36, Last annotation update)
DE   Mast cell protease-like protein precursor (EC 3.4.21.-).
GN   MCP1L.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=91107701; PubMed=1988455;
RA   Serafin W.E., Sullivan T.P., Conder G.A., Ebrahimi A., Marcham P.,
RA   Johnson S.S., Austen K.F., Reynolds D.S.;

```



CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1. TRYPTASE SUBFAMILY.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (see <http://www.isb.ch/announce/>  
CC or send an email to [license@isb.ch](mailto:license@isb.ch)).  
CC -----  
DR EMBL; U67910; AAB48263.1; -.  
DR PIR; A23698; A23698.  
DR PIR; S21275; S21275.  
DR HSSP; P20231; IAAO.  
DR MEROPS; S01.026; -.  
DR MEROPS; S01.143; -.  
DR InterPro; IPR001314; Chymotrypsin.  
DR InterPro; IPR001254; Ser\_protease\_try.  
DR Pfam; PF00089; trypsin; 1.  
DR PRINTS; PR00722; CHYMOTRYPSIN.  
DR SMART; SM00020; Tryp\_SPC; 1.  
DR PROSITE; PS50240; TRYPSIN\_DOM; 1.  
DR PROSITE; PS00134; TRYPSIN\_HIS; 1.  
DR PROSITE; PS00135; TRYPSIN\_SBR; 1.  
DR HydroLase; Serine protease; Glycoprotein; Zymogen; Signal;  
KW Multigene family.  
KW SIGNAL.  
FT SIGNAL 1 18  
FT PROPEP 19 28  
FT CHAIN 29 273  
FT ACT\_SITE 72 72  
FT ACT\_SITE 119 119  
FT ACT\_SITE 222 222  
FT ACT\_SITE 57 73  
FT DISULFID 153 228  
FT DISULFID 186 209  
FT DISULFID 218 246  
FT CARBOHYD 49 49  
FT CONFLICT 42 42  
FT CONFLICT 49 51  
SQ SEQUENCE 273 AA; 30400 MW; 65A5ED4D279FB284 CRC64;  
  
Alignment Scores:  
Pred. No.: 0.0687 Length: 273  
Score: 107.00 Matches: 46  
Percent Similarity: 37.35% Conservative: 16  
Best Local Similarity: 27.71% Mismatches: 45  
Query Match: 3.98% Indels: 58  
DB: 1 Gaps: 8  
  
US-10-020-441-1 (1-1477) x MCT7\_FAT (1-273)  
QY 91 TGNACAGSTTCACTAGTCTCAACAGAGAGAGTCAAGCTGGTCACTTGTGCTCA 150  
DB 57 CysGlyGlySerLeuIleHisProGlnTrpValLeuHisAlaHisCysVal---Gly 75  
QY 151 CCATTGCCAGTGAATCGGTAAGATGCACTGAAACACATTGTTCTCATTAATCA 210  
DB 76 ProAsnLys-----AlaAspProAsnLys 83  
QY 211 TTGATTCAACAGACAGTGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 270  
DB 84 LeuArgValGln----- 87  
QY 271 GCATATATTATATGCTATATACCTTGCTTATTTCAAGTTTCATTTTCACACAGAG 330  
DB 88 -----LeuArgLysGlnTyrLeu---TyrTyrHisAspHisLeuLeuTrpVal--- 102  
QY 331 AATGGACCAACAGACAGTCCATCAACACCGTGAAGTTAAGTGACCAAGATAC 390  
DB 103 -----SerGlnIleLeuSerHis-----ProAspPhe 111  
QY 391 ATGCCCTCTTGTATGTGGCAGCAGACAGAGACCAATGGCAGACACTCACTGATTC 450  
|||||

DB 112 Tyr-----IleAlaGln-----AspGlyAla 118  
QY 451 GATATTGCAATTGTATGCTGCTCAATGCTCACTTACAGATGAGATGAGATGATC 510  
DB 119 AspIleAlaLeuLeuLeuLeuLeuThrAsnProValAsnIleThrSerAsnValHisThrVal 138  
QY 511 AGTGTCCACAGACCATCGGATATCCCGCACCCTGGAATCGTGTGTTTCATTTGTTAT 570  
DB 139 SerLeuProProAlaSerGlnThrPheProSerGlyThrLeuCysTrpValThrGlyTrp 158  
QY 571 GGAAGCAGATGATTAACGAC 588  
DB 159 GlyAsnIleAsnAsnAsp 164  
  
RESULT 15  
MUC2\_HUMAN  
ID MUC2\_HUMAN STANDARD; PRT; 5179 AA.  
AC Q02817; Q14878;  
DT 01-JUN-1994 (Rel. 29, Created)  
DT 01-NOV-1997 (Rel. 35, Last sequence update)  
DT 15-JUN-2002 (Rel. 41, Last annotation update)  
DE Mucin 2 precursor (Intestinal mucin 2).  
GN MUC2 OR SMUC.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
OC NCBI\_Taxid=9606;  
RN [1]  
RP TISSUE=Intestine;  
RC MEDLINE=94132002; PubMed=8300571;  
RX Gum J.R. Jr., Hicks J.W., Toribara N.W., Siddiki B., Kim Y.S.;  
RT "Molecular cloning of human intestinal mucin (MUC2) cDNA.  
RT Identification of the amino terminus and overall sequence similarity  
RT to prepro-von Willebrand factor";  
RL J. Biol. Chem. 269:2440-2446(1994).  
RN [2]  
RP SEQUENCE OF 626-1895 AND 4196-5179 FROM N.A.  
RC TISSUE=Colon;  
RX MEDLINE=93016075; PubMed=1400449;  
RA Gum J.R. Jr., Hicks J.W., Toribara N.W., Roche E.-M., Lagace R.E.,  
RA Kim Y.S.;  
RT "The human MUC2 intestinal mucin has cysteine-rich subdomains located  
RT both upstream and downstream of its central repetitive region.";  
RL J. Biol. Chem. 267:21375-21383(1992).  
RN [3]  
RP SEQUENCE OF 1343-1895 AND 4176-4195 FROM N.A.  
RX MEDLINE=91358717; PubMed=1885763;  
RA Toribara N.W., Gum J.R. Jr., Culhane P.J., Lagace R.E., Hicks J.W.,  
RA Petersen G.M., Kim Y.S.;  
RT "MUC-2 human small intestinal mucin gene structure. Repeated arrays  
RT and polymorphism.";  
RL J. Clin. Invest. 88:1005-1013(1991).  
RN [4]  
RP FUNCTION: COATS THE EPITHELIA OF THE INTESTINES, AIRWAYS, AND  
RP OTHER MUCOS MEMBRANE-CONTAINING ORGANS. THOUGHT TO PROVIDE A  
RP PROTECTIVE, LUBRICATING BARRIER AGAINST PATHOGENS AND INFECTIOUS  
RP AGENTS AT MUCOSAL SURFACES.  
RN [5]  
RP SUBUNIT: MULTIMERIC.  
RN [6]  
RP SUBCELLULAR LOCATION: Secreted.  
RN [7]  
RP TISSUE SPECIFICITY: COLON, SMALL INTESTINE, COLONIC TUMORS,  
RN [8]  
RP BRONCHUS, CERVIX AND GALL BLADDER.  
RN [9]  
RP PTM: ALL CYSTEINE RESIDUES ARE INVOLVED IN INTRACHAIN OR  
RN [10]  
RP INTERCHAIN DISULFIDE BONDS (BY SIMILARITY).  
RN [11]  
RP POLYMORPHISM: THE NUMBER OF REPEATS IS HIGHLY POLYMORPHIC AND  
RN [12]  
RP VARIES AMONG DIFFERENT ALLELES.  
RN [13]  
RP SIMILARITY: THE N-TERMINAL DOMAIN SHOWS SOME SIMILARITY TO THAT  
RN [14]  
RP OF SILKMOB HEMOCYTIN.  
RN [15]  
RP SIMILARITY: CONTAINS 2 WFEC DOMAINS.  
RN [16]  
RP SIMILARITY: CONTAINS 1 C-TERMINAL CYSTINE KNOT-LIKE (CTCK) DOMAIN.  
RN [17]  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its



```

Db 1699 oSerProthThrThrProSerSerProIleThrThrThrThrThrProSerSerThrTh 1719
Oy 795 ACTGTCACAAAGTCAACCAACAGATGTTGTTATTCAGGCTGACCAATACAAAC 736
Db 1719 rThrProSerProProProThrThrMetThr-----ThrProSerProThrThr 1736
Oy 735 CGGATCAGGTAGACAGACAGACAAACAGATACACAGACACAGAAATACGACTG 676
Db 1736 rProSerProProThrThrMetThrThrLeuProProThr----- 1750
Oy 675 ACATCTGACGACGACTGATGTCGTTTATTCACCAACACTCTTCTTCATAT 616
Db 1751 -----ThrThrSerSerProLeuThrThrThrProLe 1761
Oy 615 TCCACCAATCTTACGTCAGGATCAGGTCGTTATTCATCCTTCCATACCAATGAA 556
Db 1761 uProProSerIleThrProProThrPheserProPheserThrThrProThr----- 1779
Oy 555 AACACCACTT----- 546
Db 1780 -ThrProCysValProLeuCysAsnTrpThrGlyTrpLeuAspSerGlyLysProAsn 1799
Oy 545 -----CCAGGTGGGGG-----ATATCCGATGCTGTGGCAGACTGACTCT 502
Db 1799 eHisLysProGlyGlyAspTrpGluLeuIleGlyAspValCysGlyTrpGlyTrpAla 1819
Oy 501 GATTCACACTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 442
Db 1819 aAsnIleSerCysArgAlaThrMetLysTrpAspValProIleGlyGluLeuGlyGln 1839
Oy 441 GAGTGTCTGT-----GCCATTGCTCTCTGTCTGTCTGTCTGTCTGTCTGTCT 388
Db 1839 rValValCysAspValSerValGlyLeuIleCysLysAsnGluAspGlnLys----- 1856
Oy 387 TCTGTGTGTCACCTTA 372
Db 1857 -ProGlyGlyValIle 1861

RESULT 16
MCT4_RAT STRAND: PRT; 246 AA.
ID MCT4_RAT STRAND: PRT; 246 AA.
AC P97592;
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DE Mast cell protease IV precursor (BC 3.4.21.-) (RMCP-IV) (RMCP-4).
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
RN NCBI_TaxID=10116;
RX STRAIN=Histar;
RA MEDLINE=97149430; PubMed=8996238;
RA Lutzelischwab C., Pejler G., Aveskog M., Hellman L.;
RT "Secretory granule proteases in rat mast cells. Cloning of 10
RT different serine proteases and a carboxypeptidase A from various rat
RT mast cell populations."
RL J. Exp. Med. 185:13-29(1997).
CC -1- SUBCELLULAR LOCATION: Secretory granules.
CC -1- TISSUE SPECIFICITY: MAST CELLS.
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1. GRANZYME SUBFAMILY.
CC This Swiss-Prot entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
DR EMBL; U67907; AAB8260.1; -
DR HSSP; P00770; 3RP2.

```

```

DR MEROPS; S01.005; -.
DR InterPro; IPR001314; Chymotrypsin.
DR InterPro; IPR001254; Ser_protease_Try.
DR Pfam; PF00089; trypsin; 1.
DR PRINTS; PR00722; CHYMOTRYPSIN.
DR SMART; SM00020; TRYD_SPC; 1.
DR PROSITE; PS02404; TRYPSIN_DOM; 1.
DR PROSITE; PS00134; TRYPSIN_HIS; 1.
DR PROSITE; PS00135; TRYPSIN_SER; FALSE NEG.
KM Hydrolyase; Serine protease; Signal; Zymogen; Multigene family.
FT SIGNAL 1 18
FT PROPEP 19 20
FT CHAIN 21 246
FT ACT_SITE 65 65
FT ACT_SITE 109 109
FT ACT_SITE 202 202
FT DISULFID 50 66
FT DISULFID 143 208
FT DISULFID 174 187
SQ SEQUENCE 246 AA; 27042 MW; 412F51D05098778 CRC64;

Alignment Scores:
Pred. No.: 0.126 Length: 246
Score: 104.00 Matches: 53
Percent Similarity: 35.5% Conservative: 22
Best Local Similarity: 25.12% Mismatches: 64
Query Match: 3.87% Indels: 72
DB: 1 Gaps: 10

US-10-020-441-1 (1-1477) x MCT4_RAT (1-246)
Oy 13 TTGATACGTAAGTGTGAACCTGTGACACCGCACTGAATTCCTTCATCGCATTTCTTA 72
Db 21 IleIleGlyGlyValGluSerIleProHisSerArgProGlyMetAlaLeuLeuLysIle 40
Oy 73 ACGACGAGAGA-----ACATGTGTACAGSTTCACTAGTGTCAACGAGACAGTA 123
Db 41 ValThrGluGluGlyHisValThrPheCysGlyGlyPheLeuIleSerLeuGlnPheVal 60
Oy 124 CTCACAGCTGTCATGTTGTTGTTGTCACCATGTCACCATGTCGCGTGAAGATGCACTG 183
Db 61 LeuThrAlaAlaHisCys----- 66
Oy 184 AAACACATTTGTCATCATGATGATGATGATGATGATGATGATGATGATGATGATGAT 243
Db 67 ---His-----GlyArgGluIleThrValThrLeuGlyAla----- 77
Oy 244 GTTTGTGTTGTGTATGACAGTGGGTTTGATGATTTTATATGCTATATACTTGGTCTTA 303
Db 77 ----- 77
Oy 304 TTTCAAGTTTCAATTTCTCACATGAGAAATGCGACCAACAGCATTCACCAACCG 363
Db 78 -----HisAspMetSerLysArgGluSerThrGlnGlnLysIle----- 90
Oy 364 TCTGAGTTAAGGTGCGACACAGATACATACCTCTTGTATGTCGCGACAGACAGAGAGA 423
Db 91 -----LysVal-----ValLysGlnLysPhe 97
Oy 424 CCAATGCGACAGACACTC---AGTGATTC---GATATTGCAATTGTAATGCTGGCTCA 477
Db 98 ProLeuLysTyrAsnLeuPheSerAsnPheArgAspIleMetLeuLysLeuGlnGln 117
Oy 478 ATGGTCAACTTTCAGAGTGAATCAGAGATCAGTCTCCACAGCATCGGATATCCG 537
Db 118 LysAlaValLeuThrProSerValAsnValIleProLeuProGlnHisSerLysPheIle 137
Oy 538 CCACCTGGAAGTGTGTTTTCATTTGTTGTTATGGAAGGATGATGAACCGGTGATCGG 597
Db 138 LysProGlyThrMetCysLeuAlaIleGlyTrpGlyGln---ThrGlyValLysGluPro 156
Oy 598 TCACGTAAGATGTTGGAATTTGAAGAAAGTG 630

```





AC P40209;  
 DT 01-FEB-1995 (Rel. 31, Created)  
 DT 01-FEB-1995 (Rel. 31, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE GAT2 protein.  
 GN GAT2 OR YMR136W OR YMR375.05.  
 OS Saccharomyces cerevisiae (Baker's yeast).  
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
 OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.  
 NCBI\_TaxID=4932;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=2886 / AB972;  
 RA Badcock K., Churcher C., Barrell B.G., Rajandream M.A., Walsh S.V.;  
 RL Submitted (DEC-1994) to the EMBL/GenBank/DBJ databases.  
 RN [2]  
 RP IDENTIFICATION.  
 RX MEDLINE=99320744; PubMed=10392447;  
 RA Cox K.H., Pinchak A.B., Cooper T.G.;  
 RT "Genome-wide transcriptional analysis in S. cerevisiae by mini-array  
 RT membrane hybridization";  
 RL Yeast 15:703-713(1999).  
 CC -1- SIMILARITY: CONTAINS 1 GATA-TYPE ZINC FINGER.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See http://www.isb-sib.ch/announce/  
 CC or send an email to license@isb-sib.ch).  
 CC -----  
 CC DR EMBL; 247071; CAA87350.1; -.  
 DR HSP; P17679; IGFN.  
 DR TRANSPAC; T03314; -.  
 DR SGP; S0004744; GAT2.  
 DR InterPro; IPR00679; Znf\_GATA.  
 DR Pfam; PF00320; GATA; 1.  
 DR SMART; SM00401; ZNF\_GATA; 1.  
 DR PROSITE; PS00344; GATA\_ZN\_FINGER\_1; FALSE\_NEG.  
 DR PROSITE; PS00114; GATA\_ZN\_FINGER\_2; 1.  
 KM DNA-binding; Zinc-finger.  
 FT ZN\_FING 497  
 FT SEQUENCE 560 AA; 63139 MW; 85ADECA7847208B5 CRC64;  
 Alignment Scores:  
 Pred. No.: 0.174 Length: 560  
 Score: 103.50 Matches: 72  
 Percent Similarity: 39.86% Conservative: 44  
 Best Local Similarity: 24.74% Mismatches: 81  
 Query Match: 3.88% Indels: 94  
 DB: 1 Gaps: 17  
 US-10-020-441-1 (1-1477) x GAT2\_YEAST (1-560)  
 QY 1347 AGAGGTCCACCACTGTCTGAGCGGGTAAGTCTCGGAATTCGACTGCTTTCACA 1288  
 DB 2 GlnAlaProAsnIleTyrProPheSerGlnThnGlnProGlnAlaLeuProGlyPheThr 21  
 QY 1287 CATATGCAATGCTGTCGTCGATTCGGCATTCATTAATAGTCGCTCGACCTGTGGT 1228  
 DB 22 Tyr---GlyProProGlnLeuValPheAspHisSerAlaPro-ArgValAspProLeu-- 39  
 QY 1227 GGAACAACAACACACAGACAGTACAGTACACAAAC-----GTACAGTACAGAT 1177  
 DB 40 -----His-SerThnValThrIleLysSerProLeuProLeuGlnHis---- 53  
 QY 1176 GAGAAAGAACAAACACACGGAACAAACAAACACACAGTGAATTCGAAATATATCA 1117  
 DB 54 -----TyrAsnGly-----ProAsnIleHisIleAsnSerAlaAsnAsn 67  
 QY 1116 ACTGGCAATATTCACCCCTTTACTACTACCCACACAAACATATTCACATACATCAT 1057  
 DB 1116 ACTGGCAATATTCACCCCTTTACTACTACCCACACAAACATATTCACATACATCAT 1057

DB 67 snTYR-----AlaTYRtyrTyHisIspProAsnAsnAsnAsnAsnHisSer- 84  
 QY 1056 GTGCATATATTCGACGATCAAGTCACTCTCCATCTCCACATCTCCAAACATATCTCA 997  
 DB 85 -----AsnAsnThrIleLysAsnAsnAsnIleAsnSerValLeuProAlaValAsnI 102  
 QY 996 AACCATTTGACCCCTCACTTATGATGACAAACACATTCATCGTTTCATATGATATTC 937  
 DB 102 LeGlnIleSerAsnAsnSerHisTYRArgAsnThr-----H 114  
 QY 936 ATGAG-----TCTTCCTCCCAAGGTCTCAATGCTCCATCCATCATCATTCATTC 883  
 DB 114 IsgInIleProSerAlaProGlnArgLeuValSerIleIleProAspProHisMetPro- 133  
 QY 882 CCCATTCATATCACTCACTCACTCACTCACTCACTCACTCACTCACTCACTCACTCA 823  
 DB 134 -----ProAsnIleSerHisPheGlnLeuAsnIleHisP 146  
 QY 822 CCGAGATGAAATCATCTGTCTACTGAACTGACAAAGTCAATCCAAACACAGTTGTTAT 763  
 DB 146 roGlnMet-----HisAlaProValAlaThrIspIle-His 157  
 QY 762 TATCAA-----GGCTGACCAATATCAAC 739  
 DB 158 PheGlnAlaValProValTYRAsnLysThrAsnAsnGlyIleGlyThrAspAsnIleAsn 177  
 QY 738 AACCGATCAGGTAGACACACACACACAAACAGATACACAGACACAGATTAATGTA 679  
 DB 178 Asn-----AspLysProValAsn-----SerAsnGlnAsnGlnVal 189  
 QY 678 CTGACATCT--GACTGACGACTGAGTCATGTC-----GTTTATTCACCAACACT 631  
 DB 190 LeuAspAsnIleAspGlnArgSerCysHisGlnIleAsnArgValValSerPheSerLys 209  
 QY 630 CACTTCTTCATATTCACACATCTTACGAGCAGATCAGCGGTATATATCCTTCC 571  
 DB 210 HisPhe----- 211  
 QY 570 ATACCAACAATGAAACACACAGTTC 544  
 DB 212 ---GluAsnAsnGlnLeuThrThrThr 219  
 RESULT 19  
 VU47\_HSV6U STANDARD; PRT; 651 AA.  
 AC VU47\_HSV6U  
 DT 01-OCT-1996 (Rel. 34, Created)  
 DT 01-OCT-1996 (Rel. 34, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Glycoprotein U47 precursor.  
 GN U47 OR BHLF2.  
 OS Human herpesvirus (type 6 / strain Uganda-1102) (HHV6).  
 OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;  
 OC Betaherpesvirinae; Roseolovirus.  
 NCBI\_TaxID=10370;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=93091236; PubMed=133836;  
 RA Compels U.A., Carss A.L., Sun N., Arrand J.R.;  
 RT "Infectivity determinants encoded in a conserved gene block of human  
 RT herpesvirus-6";  
 RL J Virol 67:125-131(1992).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=95266321; PubMed=7747482;  
 RA Compels U.A., Nicholas J., Lawrence G., Jones M., Thomson B.J.,  
 RA Martin M.E., Efsthaliou S., Craxton M., Macaulay H.A.;  
 RT "The DNA sequence of human herpesvirus-6: structure, coding content,  
 RT and genome evolution.";  
 RL Virology 209:29-51(1995).  
 CC -1- SIMILARITY: BELONGS TO A FAMILY THAT GROUP TOGETHER HSV-6 AND  
 CC HSV-7 U47 AND HCMV UL74.  
 CC -1- CAUTION: IT IS UNCERTAIN WHETHER MET-1 OR MET-34 IS THE INITIATOR.





```

Qy 1173 -----AAAGAACAAACACGGAACAAACAAACAAACAAAGTGAATTCGAAATATC 1118
Db 589 1CythrGluCythrGluThrGluThrSerThrSerThrProtyAlaThrSerSe 609
Qy 1117 AACTGGCAATATCCACCCCTTACATCCACACAAACAAATGTCATATCAATCA 1058
Db 609 rThrIlyThrAlaThrSer-PheThrAla-----SerThrSerAspThrM 624
Qy 1057 TGTGCAATATTCGACGATCAAGTCAAGTCAATCTC-----ACCACCTCAAA 1007
Db 624 eThrSerLeuValGlnThrAspThrThrValSerPheSerLeuSerThrThrValSerG 644
Qy 1006 CACATATCTCAACACATTCGACCCCTTACTTCAATGATACACACAACTTCATCGTTTGA 947
Db 644 1uH1eThrAspAla-----ProThrSerSerValGluSerAspAlaSerThrPhe 661
Qy 946 TAGATATTCATGATCTTCTCTCCAGTGCTCATATATCCATCATCAATCA----- 891
Db 661 1eSer-----SerAsnIyGlySerValIySerTyThrValThrSerSerI 676
Qy 890 --CATTCATC-----CCATCCAAATCTACCTCACTCACTCACTCACTCA 845
Db 676 1eH1eSerIleThrProMetTyThrProSerAsnGlnThrValThrSerSerValValS 696
Qy 844 CCTCACTCACCACCAACATGACCGAGATGGAATCACTCTGCTCACTGAAGT----- 792
Db 696 eThrProIleThrSerGluSerSerGluSerSerAlaSerValThrIleLeuProSerT 716
Qy 791 -----TCACAAAGTCAATCCAAACACAGTTGTTATTCAGAGCTTGACCAATACA 740
Db 716 h1eThrSerGluThrPheIySerProSerThrMet-----LysThrIySerVal 731
Qy 739 CAACGGATCGAGTGAAGACAGACACACAAACAGATACACAGACACAGATACAGT 680
Db 731 a1SerIleSerSerSerProThrAsnIleThrSerTyThrAspThr-ThrSerIyAsp 750
Qy 679 ACTGACATCTGACGACGATGATGATGATGATGATGATGATGATGATGATGATGAT 620
Db 751 SerThrValGlySerSerThrSerSerValSerIleSerSerIleSerLeuProSer 770
Qy 619 -----ATATTCACCAATCTTACGTCAGATGATGATGATGATGATGATGAT 584
Db 771 SerTySerAlaSerSerGluGlnIlePheH1eSerSerIleValSerSerAsnGlyGln 790
Qy 583 TATCATCCCTTCATTAACCAACATGAACAAACACAGTTCCAGTGGCGGATATCCGATG 524
Db 791 -AlaLeuThrSerPheSerSerThrIyValSerSerSerGluSerSerGluSerH1eAs 810
Qy 523 GCTGGGACGATGATCACTGATTCATCTGATGATGATGATGATGATGATGATGATGAT 464
Db 810 gThrSerProThrThrThrSerSerGluSerGlyIle-----LysSerSe 824
Qy 463 CAATTGCAATTCGAATCCACT 442
Db 824 rGlyValGlnIleGluSerThr 831

RESULT 22
MCT9_MOUSE STANDARD; PRT; 246 AA.
AC 03164;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 15-JUL-1998 (Rel. 36, Last annotation update)
DE Mast cell protease 9 precursor (BC 3.4.21.-) (MCP-9).
GN MCP9.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BALB/CJ;
RX MEDLINE=98030598; PubMed=9360993;

```

```

RA Hunt J.E., Friend D.S., Gurish M.F., Feyfant E., Sali A., Huang C.,
RA Ghildyal N., Stechschulte S., Austen K.F., Stevens R.L.;
RT "Mouse mast cell protease 9, a novel member of the chromosome 14
RT family of serine proteases that is selectively expressed in uterine
RT mast cells."
RL J. Biol. Chem. 272:29158-29166(1997).
CC -1- TISSUE SPECIFICITY: SELECTIVELY EXPRESSED IN UTERINE MAST CELLS.
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1. GRANZYME SUBFAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; AF007119; AAB65245.1; -.
CC EMBL; AF007120; AAB65246.1; -.
CC HSSP; P00770; JRP2.
CC MGP; MGI:1194491; MCP9.
CC InterPro; IPR001314; Chymotrypsin.
CC InterPro; IPR001254; Ser_protease_Try.
CC P1am; PF00089; Trypsin; 1.
CC PRINTS; PR00722; CHYMOTRYPSIN.
CC SMART; SM00020; Tryp_Spc; 1.
CC PROSITE; PS00240; TRYPSIN_DOM; 1.
CC PROSITE; PS00134; TRYPSIN_HIS; 1.
CC PROSITE; PS00135; TRYPSIN_SER; 1.
CC KW Hydroxylase; Serine protease; Zymogen; Signal; Multigene family.
CC FT SIGNAL 1 18 BY SIMILARITY.
CC FT PROPEP 19 20 ACTIVATION PEPTIDE (BY SIMILARITY).
CC FT CHAIN 21 246 MAST CELL PROTEASE 9.
CC FT ACT_SITE 65 65 CHARGE RELAY SYSTEM (BY SIMILARITY).
CC FT ACT_SITE 109 109 CHARGE RELAY SYSTEM (BY SIMILARITY).
CC FT ACT_SITE 202 202 CHARGE RELAY SYSTEM (BY SIMILARITY).
CC FT DISULFID 50 66 BY SIMILARITY.
CC FT DISULFID 143 208 BY SIMILARITY.
CC FT DISULFID 174 187 BY SIMILARITY.
CC SQ SEQUENCE 246 AA; 26652 MW; C3BFC5D7B5DD01 CRC64;

Alignment Scores:
Pred. No.: 0.174 Length: 246
Score: 102.50 Matches: 42
Percent Similarity: 32.02% Conservative: 23
Best Local Similarity: 20.69% Mismatches: 51
Query Match: 3.82% Indels: 87
DB: 1 Gaps: 7

US-10-020-441-1 (1-1477) x MCT9_MOUSE (1-246)
Qy 55 CCGTTCATCGCATCTTCTTAACGACGACGAG-----ACATGTGTCAGAST 99
Db 33 ProTyMetAlaTyValAsnThrPheSerIySylGlyTyValAlaIleCySeGlyGly 52
Qy 100 TCATAGTCTCAAGACGAGACGACTCACTGAGTGTGATG-----GTTGTC 147
Db 53 PheLeuIleAlaProGlnPheValMetThrAlaAlaIleCySeSerGlyArgArgMetThr 72
Qy 148 TCACATTCGACGATTCGGGTGAGAGATGAGATGAGTGAACACATGTTGTCATATGAT 207
Db 73 ValThrIleuGlyAlaH1eAsnValAlaGlyArg-----GluCySerThr 86
Qy 208 CGATTGATTTACACAGACAGGTGTTGCGTGCGTGCGTTGTTGTTGTTGTTGATGAGTGG 267
Db 87 Gln----- 87
Qy 268 TTTGATGATATTTAATGTATATATACGTGCTTATTTCAGTTTCAATTCACATG 327
Db 87 ----- 87
Qy 328 AGGATGGCGACCAACAGGATCATCAACACCGTTCGAGTTAAGTGGACACAGGA 387
Db 87 ----- 87

```

D8 88 -----GlnylsIleValGluLysTyr 95

QY 388 TACATGCCCTCTTGTAAGTCGGACGACAGAGAGACAACAAATCGCACACTCACTAGTGA 447  
:::|  
Db 96 IleuPhePro-----ProSerTyrAsnValSerSerLys 106

QY 448 TTC---GATTATTGCATTTGTATGCTGGCTCAAAATGTCATCTTAACAGTGGATCAGA 504  
||| ||||| ::::: ||| ::: ||||| ||| :::  
Db 107 PheaaNapIleValLeuIleuLysLysLysGlnAlaAsnLeuThrSerAlaLasp 126

QY 505 GTGATCACTCTGCCACACACCATTCGATATCCCGCCACCTGGAACTGATGTTTCACTTGT 564  
||||: |::: ||||| ||||| ||||| ||||| :::  
Db 127 ValValPProLeuPProGlyPProSerAspHeaLalysPProGlyThrMetCysTrpAlaLa 146

QY 565 GGTATVGAAGGATGATPAACGACCGTGATCCGTACCGTAAGATGTTGGAAATTATYGAG 624  
|||:|:| ||||| ||||| ||||| ||||| |||||  
Db 147 GlyTPglYArgThr-----GlyValLysLys 155

QY 625 AAAGTAGCT 633  
:::|  
Db 156 SerIleSer 158

RESULT 23

CH13 CANAL STANDARD: PRT: 567 AA.

ID CH13 CANAL STANDARD: PRT: 567 AA.

AC P40954;  
DT 01-FEB-1995 (Rel. 31, Created)  
DT 01-NOV-1995 (Rel. 32, Last sequence update)  
DT 01-NOV-1995 (Rel. 32, Last annotation update)  
DE Chitinase 3 precursor (EC 3.2.1.14).  
GN CHT3.  
OS Candida albicans (Yeast).  
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
OC Saccharomycetales; mitosporic Saccharomycetales; Candida.  
RX NCBI\_TaxID=5476;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=ATCC 10261;  
RZ MEDLINE=95223977; PubMed=7708682;  
RA McCreath K.J., Specht C.A., Robbins P.W.;  
RL "Molecular cloning and characterization of chitinase genes from  
RT Candida albicans".  
RT Proc. Natl. Acad. Sci. U.S.A. 92:2544-2548(1995).  
CC -|- CATALYTIC ACTIVITY: Hydrolysis of the 1,4-beta-linkages of N-acetyl-D-glucosamine polymers of chitin.  
CC -|- SUBCELLULAR LOCATION: Secreted (Probable).  
CC -|- SIMILARITY: BELONGS TO CHITINASE CLASS II (FAMILY 18 OF GLYCOSYL HYDROLASES).

-----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see <http://www.isb-sib.ch/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

-----

DR EMBL: U15801; AAA68016.1; -.  
DR HSSP: P23472; ZHYM.  
DR InterPro: IPR001579; Chitinase\_18/2.  
DR InterPro: IPR001223; Glyco\_hydro\_18.  
DR Pfam: PF00704; Glyco\_hydro\_18; 1.  
DR PROSITE: PS01095; CHITINASE\_18; 1.  
KW Hydrolase; Glycosidase; Chitin degradation; signal; Glycoprotein;  
KW Chitin-binding; Multigene family.  
FT SIGNAL 1..16 POTENTIAL.  
FT CHAIN 17..567 CHITINASE 3.  
FT ACT SITE 157..157 PROTON DONOR (BY SIMILARITY).  
FT CARBOHYD 159..159 N-LINKED (GLCNAC... ) (POTENTIAL).  
FT DOMAIN 319..436 SSR/TRH-RICH.  
SQ SEQUENCE 567 AA, 60060 MW, DB043126F5E22C2 CRC64;

Alignment Scores:

Pred. No.:	0.216	Length:	567
Score:	102.50	Matches:	65
Percent Similarity:	39.06%	Conservative:	35
Best Local Similarity:	25.39%	Mismatches:	89
Query Match:	3.84%	Indels:	67
DB:	1	Gaps:	8

US-10-020-441-1 (1-1477) x CH13\_CANAL (1-567)

QY	1219	ACAACCAACAACAAGTACATGTCACAAACGTCACATGTCAGTAAGAAGAACAAAACAC	1160
Dd	325	ThrThrThrThrThr-----ThrSerThrThrIleSerSerSerSerSerSerSerSerTh	344
QY	1159	AACGAACAACAAAACACACAAGGAAATTGGAAAATATACCTGGCACAATATCCAC	1100
Dd	344	IserLythrSerThrThrThrThrThrThrThrSerSerSerIleSerSerThrThrSerSerTh	364
QY	1099	CCTTTACATACCACACACAATAAGTCACATACATCACATGTGCACCTAATATTGCCAG	1040
Dd	364	RThSerSerThr-----	368
QY	1039	ATCAAGTACTCTTCATCATCTGCACCATCTCTCCAACAACATATCTCAAACATTGACCCCCTA	980
Dd	369	SerSerSerSerSerSerSerSerSerSerSerSerSerThrThrSerSerSerThrThrSe	388
QY	979	CTTATGTACACAACACATCTTCATCGTTGGCTAGATATTTCATGAGCTTCTCTCTC	920
Dd	388	RglIleSerThrThr-----SerThrAlaproThse	399
QY	919	CAAGGGCTCAATCATCCCATCATACATCCATCCATCATCATCTCACTCACTCA	860
Dd	399	rSerThrSerIleSerSerSerSerThrIleSerThrSerAlaserThrSerAspThrThSe	419
QY	859	CCTCACATCAGTCACCTCACCTCACCAACACAAACGAGATGGAAATCACTCTGCTA	800
Dd	419	rValThrSerSerglu---ThrThrPro-----	427
QY	799	CTGAACGTCTCAAAAGTCACAAACACAGCTTTTATTCAAAGGCTCTGCACATATACA	740
Dd	428	----ValValThrProSerSerIleSerSerAlalThrThrIleProGlyAsp-SerThrT	446
QY	739	CAACCGGATCAGTAGACAGACAGACACAAACAGATACACAGACA-----	693
Dd	446	hTrnGlyIleSerLyseSerSerSerThrLyseProAlaThrSerThrThrSerAlaleus	466
QY	692	-----CACAGAAATACAGCTGATGATCTGCATGACGACATGATGATGCTTATATACC	638
Dd	466	erSerSerThrThrThrValAlaThrIleProasp--LysGluIleIleAsnThrT	485
QY	637	AACAACATCATTTCTTCAATATATCCACCATTTCTTAGCGTAGCGATCGCTTATAT	578
Dd	485	OThr-----AspThrGluThrThrSerLyseProProAlaIleIleTh	499
QY	577	CCCCCTCC-----ATAACCAACATGAAGAACACACAGTTCACAGGTGGCGGATAT	530
Dd	499	RglSeraspAlaThrThrThrIleThrGlnAsnLeuthrProSerThrThrThrLysAsnVa	519
QY	529	CCGA-----TGGCTGTGG	517
Dd	519	IlyThrThrSerThrThrAsnIleValThrGluThrPValITrp	532

RESULT 24  
MCT2\_MOUSE  
ID MCT2\_MOUSE STANDARD; PRT; 244 AA.

DT 01-APR-1990 (Rel. 14, Created)  
DT 01-OCT-1996 (Rel. 34, Last sequence update)  
DT 15-JUL-1998 (Rel. 36, last annotation update)  
DE Mast cell protease 2 precursor (EC 3.4.21.-) (MMCP-2).  
GN MCP2.  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.







```
CC -1- CATALYTIC ACTIVITY: Protein tyrosine phosphate + H(2)O = protein
CC tyrosine + phosphate.
CC -1- SUBCELLULAR LOCATION: Cytoplasmic.
CC -1- TISSUE SPECIFICITY: IN THE ANTERIOR-LIKE AND PRETALK CELL TYPES.
CC -1- DEVELOPMENTAL STAGE: EXPRESSED AT MODERATE LEVELS DURING GROWTH
CC AND DEVELOPMENT.
CC -1- SIMILARITY: BELONGS TO THE NON-RECEPTOR CLASS OF THE PROTEIN-
CC TYROSINE PHOSPHATASE FAMILY.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.ebi.ac.uk/announcements or send an email to license@ebi.ac.uk).
CC -----
DR EMBL; U38197; AAC47041.1; -.
DR HSSP; O06124; 2SHP.
DR DictyDb; DD07777; ptpC2.
DR InterPro; IPR000387; TYR_phosphatase.
DR InterPro; IPR000242; TYR_Pp.
DR Pfam; PF00102; Y_phosphatase; 1.
DR PRINTS; PR00700; PRTYPHPTASE.
DR SMART; SM00194; PTPc; 1.
DR PROSITE; PS00383; TYR_PHOSPHATASE_1; 1.
DR PROSITE; PS50056; TYR_PHOSPHATASE_2; 1.
DR DR PROSITE; PS50055; TYR_PHOSPHATASE_PTP; 1.
KW Hydrolyase.
FT ACT SITE 649 649 BY SIMILARITY.
FT DOMAIN 460 716 PROTEIN-TYROSINE PHOSPHATASE.
FT DOMAIN 64 71 POLY-ASN.
FT DOMAIN 109 118 POLY-ASN.
FT DOMAIN 137 190 POLY-ASN.
FT DOMAIN 249 257 POLY-SER.
FT DOMAIN 268 265 POLY-THR.
FT DOMAIN 286 289 POLY-ASN.
FT DOMAIN 366 371 POLY-SER.
FT DOMAIN 787 790 POLY-GLN.
FT DOMAIN 834 839 POLY-GLN.
FT DOMAIN 883 892 POLY-GLN.
FT DOMAIN 906 914 POLY-ASN.
FT DOMAIN 943 963 POLY-ASN.
SQ SEQUENCE 989 AA; 109995 MW; 9371105AFB0974AF CRC64;

Alignment Scores:
Pred. No.: 0.308 Length: 989
Score: 101.50 Matches: 62
Percent Similarity: 32.00% Conservative: 34
Best Local Similarity: 20.67% Mismatches: 74
Query Match: 3.81% Indels: 130
Gaps: 9

DS-10-020-441-1 (1-1477) x PTPJ_DICDI (1-989)
QY 1220 AACACCAACAACAAGTAGTAAGTACACAAAGTACAGATGAGAAGAACAAAACA 1161
Dd 165 AenAenAenAenAenAenAenAenAenAenAenAenAenAenAenAenAenAenA 184
QY 1160 CAACGAACAACAAAACAACAACAAGTAGTAATTGGAAAAATATCAACTGGCACAA 1106
Dd 184 nAsnAenSerAenSerAenAenIleGlulIeAsnValProSerIleGlnPheAspAsnGluPr 204
QY 1105 -----ATCCACCCTTTTACACACACCAACAACAACAATATGCCACCATC 1065
Dd 204 cAlAMetGluValApSerValAlAlProLeuAsnValProSerAen-----HisTh 221
QY 1064 AATCAATGTGCATCATATATTCGACAGCATCAAGTCACTTCACATCGACCA----- 1016
Dd 221 rArghrThrIeuAlaMetHisAenThrIylSerIeuSerThrIrrSerAenIleGlyLeuLe 241
QY 1015 -----CTCTCCAAC----- 1006
```

```

Db      241 uasnlleue-ProhanglinserserserserserThrThrT 265
           ||| |||||
Oy      1005 -----ACACTCAAAACATTGACCCTTACTTCAG- -T 977
Db      261 hrThrThrThrSerSerSerLeuWetProGlnSerLeuPheAAsnSerThrt 281
Oy      971 ACACAACCACTTC-----
Db      281 yrasnaashihiaasnaasnasnsersernaalaglyilevalglyeuaeng 301
Oy      958 -----
Db      301 lyserThSerSerLeuProThrGlnalaglnvalglnleuglnlmelglnlmm 322
Oy      957 -----CATGGTTTGATGATATTTCAATGAGCTTCTCTCCGAAGGTGTCAT 906
Db      321 etglnglnhiaglmglnhiaglnlrylsylsalaaenleuSerSerThervalv 341
Oy      905 CATCCCATCACATCATTCCATCCCATCCATCTACTCTCACTCACTCACTCAGTC 844
Db      341 alaspasasnlleaasnaasnaasnprometan-----Thiser-Thr 354
Oy      845 ACGTCACTCAACCAACAATGACCGAGATGAATCACTGCTACTGAACTGCACAA 786
Db      335 SerSerProAlaGlnProAsnalSerProPheserPheserSerSerLeuPhSer 378
Oy      785 AGTCATCCMAACAGAGTTGTTATTATCAAGGCTTGACCAATACMACCGATCAGT 726
Db      375 Asn-----SerSerleuSerAsnSerIyserylSerAla 388
Oy      725 AGACAGACAGACACAAACAGATACACAGACACACAGAATACTGACTGACT 666
Db      387 SerThrTrSerThrSerThSerSer-----
Oy      665 GACGACGTGACGTGCTGTTATTCCACCAACTCACTTCTTCATPATTCACCAATTC 606
Db      395 -----
Oy      605 TTACGTGACGATCAGGTCGTTATTCATCCCTTCATACCAACATGAAAACA 552
Db      396 -----SerAsnSerMetSerSerSerProProSerleuLystr 409

RESULT 27
NCRI HUMAN STANDARD; PRT; 2440 AA.
ID NCRI_HUMAN
AC 075376; O9UPV5; Q9U018;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DE 16-OCT-2001 (Rel. 40, Last annotation update)
DE Nuclear receptor co-repressor 1 (N-COR1) (N-COR).
GN NCOR1 OR KIA1047.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_Taxid=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Petal brain;
RX MEDLINE=98393736; PubMed=9724795;
RA Wang Y., Hoshino T., Redner R.L., Kajigaya S., Liu J.M.;
RT "ERO, fusion partner in t(8;21) acute myeloid leukemia, represses
RL transcription by interaction with the human N-COR/msin1/HDAC1
complex."
RL Proc. Natl. Acad. Sci. U.S.A. 95:10860-10865(1998).
RN [2]
RP SEQUENCE OF 782-2440 FROM N.A.
RC TISSUE=Brain;
RX MEDLINE=99397452; PubMed=10470851;
RA Kikuno R., Nagase T., Ishikawa K.-I., Hirosewa M., Miyajima N.,
RA Tanaka A., Kotani H., Nomura N., Ohara O.;
RT "Prediction of the coding sequences of unidentified human genes. XIV.
RT The complete sequences of 100 new cDNA clones from brain which code

```

for large proteins in vitro.";  
 RT DNA Res. 6:197-205(1999).  
 RN [3]  
 RP SEQUENCE OF 974-2440 FROM N.A.  
 RA Nagaya T., Chen K.-S., Fujieda M., Ohmori S., Richer J.K.,  
 RA Horwitz K.B., Lupski J.R., Seo H.;  
 RT "Localization of the human nuclear receptor co-repressor (hn-Cor) gene  
 RT between the CMT1A and the SMS critical regions of chromosome  
 RT 17p11.2.";  
 RL Genomics 59:339-341(1999).  
 CC -1- FUNCTION: MEDIATES THE TRANSCRIPTIONAL REPRESSION ACTIVITY OF SOME  
 CC NUCLEAR RECEPTORS BY PROMOTING CHROMATIN CONDENSATION, THUS  
 CC PREVENTING ACCESS OF THE BASAL TRANSCRIPTION.  
 CC -1- SUBUNIT: FORMS A LARGE COREPRESSOR COMPLEX THAT CONTAINS SIN3A/B  
 CC AND HISTONE DEACETYLASES HDAC1 AND HDAC2. THIS COMPLEX ASSOCIATES  
 CC WITH THE THYROID (TR) AND THE RETINOIC ACID RECEPTORS (RAR) IN THE  
 CC ABSENCE OF LIGAND.  
 CC -1- SUBCELLULAR LOCATION: Nuclear (By similarity).  
 CC -1- DOMAIN: THE N-TERMINAL REGION CONTAINS REPRESSION FUNCTIONS THAT  
 CC ARE DIVIDED INTO THREE INDEPENDANT REPRESSION DOMAINS (RD1, RD2  
 CC AND RD3). THE C-TERMINAL REGION CONTAINS THE NUCLEAR RECEPTOR-  
 CC INTERACTING DOMAINS THAT ARE DIVIDED IN TWO SEPARATE INTERACTION  
 CC DOMAINS (ID1 AND ID2).  
 CC -1- DOMAIN: THE TWO INTERACTION DOMAINS (ID) CONTAIN A CONSERVED  
 CC SEQUENCE REFERRED TO AS THE CORNR BOX. THIS MOTIF IS REQUIRED AND  
 CC SUFFICIENT TO PERMIT BINDING TO UNLIGANDED TR AND RARs. SEQUENCES  
 CC FLANKING THE CORNR BOX DETERMINE NUCLEAR HORMONE RECEPTOR  
 CC SPECIFICITY.  
 CC -1- SIMILARITY: CONTAINS 1 SANT-A DOMAIN.  
 CC -1- SIMILARITY: CONTAINS 1 MYB-LIKE DOMAIN.  
 CC -1- SIMILARITY: CONTAINS 2 CORNR BOXES.  
 CC -1- SIMILARITY: BELONGS TO THE N-COR NUCLEAR RECEPTOR COREPRESSORS  
 CC FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL: AF044209; AAC33550.1; -  
 CC EMBL: AB028970; BAA82999.1; -  
 CC EMBL: AB019524; BAA75814.1; -  
 CC TRANSPAC: T04687; -  
 CC GeneW: HGNC:7672; NCOR1.  
 DR MIM: 600849; -  
 DR InterPro: IPR001005; MYB DNA binding.  
 DR Pfam: PF00249; myb DNA-binding; 2.  
 DR SMART: SS00395; SANT; 2.  
 DR PROSITE: PS00900; MYB\_3; 1.  
 KM Nuclear protein; transcription regulation; DNA-binding; Repressor;  
 KM Coiled coil.  
 FT DOMAIN 174 216  
 FT DOMAIN 254 312 COILED COIL (POTENTIAL).  
 FT DOMAIN 299 328 INTERACTION WITH SIN3A/B.  
 FT DNA\_BIND 437 482 COILED COIL (POTENTIAL).  
 FT DNA\_BIND 620 670 SANT-A (POTENTIAL).  
 FT DOMAIN 501 557 MYB.  
 FT DOMAIN 607 617 COILED COIL (POTENTIAL).  
 FT DOMAIN 988 1816 PRO-RICH.  
 FT DOMAIN 2055 2059 INTERACTION WITH ETO.  
 FT DOMAIN 2263 2267 CORNR BOX OF ID1.  
 FT DOMAIN 58 64 CORNR BOX OF ID2.  
 FT DOMAIN 593 603 POLY-GLN.  
 FT DOMAIN 1032 1035 POLY-ALA.  
 FT DOMAIN 1707 1712 POLY-PRO.  
 FT DOMAIN 1952 1963 POLY-ALA.  
 FT CONFLICT 1014 1014 L->V (IN REF. 2).  
 FT CONFLICT 1508 1509 FP->SS (IN REF. 2).  
 FT CONFLICT 1561 1561 W->R (IN REF. 2).

FT CONFLICT 1567 1567 Q->H (IN REF. 2).  
 SQ SEQUENCE 2440 AA; 270263 MW; 60A4D7964D00EDAB CRC64;  
 Alignment Scores:  
 Pred. No.: 0.482  
 Score: 100.50  
 Percent Similarity: 32.32%  
 Best Local Similarity: 21.08%  
 Query Match: 3.77%  
 DB: 1  
 Gaps: 21  
 US-10-020-441-1 (1-1477) x NCRI\_HUMAN (1-2440)  
 QY 1425 ACAATGATATCGGAGAGTTGGCAGTGTGACACCATGTATGACGACCGATGCTGGA 1366  
 |||  
 |||  
 Db 1290 ThrValLeuSerGlySerIleMetClnGlyThrProAlaGlaThrThGluSerPhe--- 1308  
 QY 1365 CTTGAGGATGGG-----AGAGAGGTCACCACTGTCA 1330  
 |||  
 |||  
 Db 1309 -----GluAspGlyLeuIleTyrProIySglnIleIlyAsrGluSerProProIleArg 1326  
 QY 1329 CCGTGGAGGGGAACTGTCGCAAAATTCGACT-----GCTTCACACAT 1285  
 |||  
 |||  
 Db 1327 AlaPheGluGlyAlaIleThrIlySgIlyProTyrAspGlyIleThrIleIlySgIu 1346  
 QY 1284 ATAGAGTTGCCATGTGGTCGTCATGCGCATTCATATAGTCGTGACCCGTGTGCA 1225  
 |||  
 |||  
 Db 1347 MetGlyAsrSerIleHisGlnIleProArgGlnAspIleLeuThrGlnIu-----Ser 1364  
 QY 1224 AACAAACACCAACAAACACAGTACGTACA-----ACAAAGTAC 1186  
 |||  
 |||  
 Db 1365 ArgIlyThrProGluValAlaGlnSerThrArgProIleIleGlnIlySerIleSergIn 1384  
 QY 1185 AGTACAGATGAGAAAGAACAAACACGAGCAAAACAAACACACACAGTGAATTCG 1126  
 |||  
 |||  
 Db 1385 GlyThrProIleIlyPheAspIleHisSerGlnSerAlaIleIlyHisAsnValIlyS 1404  
 QY 1125 AAAATATCACTGCGACACATA----- 1105  
 |||  
 |||  
 Db 1405 SerLeuIleThrGlyProSerIlyLeuSerArgIlyMetProIleuGlnIleValPro 1424  
 QY 1104 -----TCCACCTT 1096  
 |||  
 |||  
 Db 1425 GluAsnIleIlyValAlaGlnArgIlyIlyTyrGluAspValIlyAlaGlyIuThrVal 1444  
 QY 1095 TACACTACCCACACACATAGTCCAC-----ATCAATACATGTCGAC 1051  
 |||  
 |||  
 Db 1445 ArgSerArgHisThrSerValIleSerGergIlyProSerValIleuArgThrIleHis 1464  
 QY 1050 TATATTTCAGCATCAAGTCACTCTCCA-----TCTGACCACTCT 1012  
 |||  
 |||  
 Db 1465 GluAlaProIyAlaGlnLeuSerProGlyIleTyrAspAspThrSerAlaArgThr 1484  
 QY 1011 CCAAAACACTACTCAAAACACTG-----ACCCCTACTTATGATGACCAACACAC 961  
 |||  
 |||  
 Db 1485 ProValSerTyrGlnAspThrMetSerArgIlySerPro-----Met-MetAsnArgTh 1502  
 QY 960 TTCACATCGTTTCATAGATATTTCATGAGTCTTCCTCTCCCAAGTGGTCAATCAT-- 903  
 |||  
 |||  
 Db 1502 rSerAspValThrIle-----ProIleAsnIlySerThrAsnHisel 1516  
 QY 902 -----CCCATCACATCACTTCCATCCCATCTCACT-- 867  
 |||  
 |||  
 Db 1516 uArgIlySerThrIleuThrProThrGlnArgIleuSerIleProAlaIlySerProValPr 1536  
 QY 866 -----CACTGACCTCACTACGTCACCTCACTCAACCAACATGACCGGAGA 817  
 |||  
 |||  
 Db 1536 cGlyValAspProValIleSerThrProPheAspProHisIleArgGlySerThrAl 1556  
 QY 816 TGGG-----AATCACTTGCTAGTCACTGACGTGACAAAGATGCTCA----- 777  
 |||  
 |||  
 Db 1556 aGlyIleuValIlyTyrPseHisIleuProThrGlnIleuAspProAlaMetProPheHisAr 1576





CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (see <http://www.ebi.ac.uk/announce/>  
 CC or send an email to [license@ebi.ac.uk](mailto:license@ebi.ac.uk)).

DR EMBL; D38455; BAA07486.1; -  
 DR EMBL; U67909; AAB48262.1; -  
 DR HSSP; P20231; IAAO.  
 DR MEROPS; S01.025; -  
 DR InterPro; IPR001314; Chymotrypsin.  
 DR InterPro; IPR001254; Ser\_protease\_Try.  
 DR Pfam; PF00089; trypsin; 1.  
 DR PRINTS; PR00722; CHYMOTRYPSIN.  
 DR SMART; SM00020; Tryp\_Spc; 1.  
 DR PROSITE; PS50240; TRYPSIN\_DOM; 1.  
 DR PROSITE; PS00134; TRYPSIN\_HIS; 1.  
 DR PROSITE; PS00135; TRYPSIN\_SER; 1.  
 KM Hydrolyase; Serine protease; Glycoprotein; Multigene family;  
 KM Signal.  
 FT SIGNAL. 1 19  
 FT PROPEP 20 29  
 FT CHAIN 30 274  
 FT ACT\_SITE 73 73  
 FT ACT\_SITE 120 120  
 FT ACT\_SITE 223 223  
 FT DISULFID 58 74  
 FT DISULFID 154 229  
 FT DISULFID 187 210  
 FT DISULFID 219 247  
 FT CARBOHYD 104 104  
 FT CARBOHYD 131 131  
 FT CONFLICT 128 139  
 FT CONFLICT 139 139  
 SQ SEQUENCE 274 AA; 30508 MW; DF84D55668CA1A25 CRC64;

Alignment Scores:  
 Pred. No.: 0.304  
 Score: 100.00  
 Percent Similarity: 35.18%  
 Best Local Similarity: 22.11%  
 Query Match: 3.72%  
 Gaps: 1 6

US-10-020-441-1 (1-1477) x MCT6\_RAT (1-274)

QY 4 TCAACCTGGTGAATGATGAGTGAACCTGTCACACCGCATGATTCCTGTCATC 63  
 DB 39 SerIysrtp-----ProTprglnValSerLeuArgPheLysPheSer 52  
 QY 64 GCATTCCTTAAGACAGAGAACATGCTGACAGTTCACCTGCTTCACAGACAGATG 123  
 DB 53 PheTprMet-----HisPheCysGlyGlnSerLeuIleHisProIleTprVal 68  
 QY 124 CTCACAGCTGGTCAATGCTGTT-----TGCCTACCATTCGACAGTGAATGGGGA 171  
 DB 69 LeuThrAlaAlaHisCysValGlyLeuHisIleLysSerPro---GluLeuPheArgVal 87  
 QY 172 AGAGATGACATGAAACCATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 231  
 DB 88 GlnLeuArgGlnGlnIleTyrLeuTyr----- 95  
 QY 232 GCGTGTGGCGGCTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 291  
 DB 96 -----TyrAlaAspGlnLeuLeuThrVal 103  
 QY 292 TACTGTGCTTAATTCAGAGTTCATTCCTCACTGACAGGATGGGACCAAGACATC 351  
 DB 104 AsnArgThrValValHisProHisIleTyrThrValGluAsp----- 117  
 QY 352 CATCACCAACCGCTGAGATTAAAGGTGACCAAGATACATGCGCTTGTATGTGGCA 411

DB 117 ----- 117  
 QY 412 CGACAGAGAGACCAATGCACAGACACTGATGATTCGATATGATTCATGCTG 471  
 DB 118 -----GlyAlaAspIleAlaLeuLeuIleu 126  
 QY 472 GCTCAATGCTCACTTAAGATGAGATGATGATGATGATGATGATGATGATGAT 531  
 DB 127 GluIleProValAsnValSerThrHisIleHisPheIleSerLeuProProIleAspGlu 146  
 QY 532 ATCCCGCACCTGCAACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 588  
 DB 147 ThrPheProSerGlyThrSerCysThrValThrGlyIleGlyAspIleAspSer 165

RESULT 31  
 TRYX\_HUMAN  
 ID TRYX\_HUMAN STANDARD; PRT; 321 AA.  
 AC Q5NRR2; Q5NRR2; Q5NRR2; Q5NRR2;  
 DT 16-OCT-2001 (Rel. 40, Created)  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Trypsin gamma precursor (EC 3.4.21.-) (Transmembrane trypsinase).  
 GN Trypsin OR TMT.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.  
 OC NCBI\_Taxid=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A. (VARIANTS GAMMA-1 AND GAMMA-2).  
 RX MEDLINE=20302813; PubMed=10843716;  
 RA Caughey G.H., Raymond W.W., Blount J.L., Hau L.W., Pallero M.,  
 RA Wolters P.J., Vergese G.M.;  
 RT "Characterization of human gamma-trypsinase, novel members of the  
 RT chromosome 15p mast cell trypsinase and procathepsin gene families";  
 RL J. Immunol. 164:6566-6575(2000).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=99452974; PubMed=10521469;  
 RA Wong G.W., Tang Y., Feyfant E., Sali A., Li L., Li Y., Huang C.,  
 RA Friend D.S., Killis S.A., Stevens R.L.;  
 RT "Identification of a new member of the trypsinase family of mouse and  
 RT human mast cell proteases which possesses a novel COOH-terminal  
 RT hydrophobic extension";  
 RL J. Biol. Chem. 274:30784-30793(1999).  
 RN [3]  
 RP SEQUENCE OF 220-321 FROM N.A.  
 RA Miltman S., Agnew W.S.;  
 RT "Organization and alternative splicing of CACNA1H";  
 RL Submitted (JAN-2001) to the EMBL/GenBank/DBJ databases.  
 CC -1- SUBCELLULAR LOCATION: MEMBRANE-ANCHORED (potential).  
 CC -1- TISSUE SPECIFICITY: Expressed in many tissues.  
 CC -1- POLYMORPHISM: There are two alleles, gamma-I and gamma-II which  
 CC differ by 5 residues.  
 CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1. TRYPSIN SUBFAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (see <http://www.ebi.ac.uk/announce/>  
 CC or send an email to [license@ebi.ac.uk](mailto:license@ebi.ac.uk)).

DR EMBL; AF191031; AAF76457.1; -  
 DR EMBL; AF195508; AAF76458.1; -  
 DR EMBL; AF175759; AAF03697.1; -  
 DR EMBL; AF175522; AAF03695.1; -  
 DR EMBL; AF223563; AAG48852.2; -  
 DR HSSP; P00763; IDPO.  
 DR MEROPS; S01.028; -  
 DR Genew; HGNC:14134; TPST1.  
 DR InterPro; IPR001314; Chymotrypsin.  
 DR InterPro; IPR001254; Ser\_protease\_Try.

DR Pfam; PF00089; trypsin; 1.  
 DR PRINTS; PR00722; CHYMOTRYPSIN.  
 DR SMART; SM00020; Tryp\_Spc; 1.  
 DR PROSITE; PS50240; TRYPSIN\_DOM; 1.  
 DR PROSITE; PS00134; TRYPSIN\_HIS; 1.  
 DR PROSITE; PS00135; TRYPSIN\_SER; FALSE NEG.  
 KM Hydroxylase; Serine protease; Signal; Glycoprotein; Zymogen;  
 KM Transmembrane; Polymorphism.  
 FT SIGNAL 1 19 POTENTIAL.  
 FT CHAIN 20 36 TRYPTASE GAMMA LIGHT CHAIN.  
 FT CHAIN 38 321 TRYPTASE GAMMA HEAVY CHAIN.  
 FT TRANSMEM 284 304 POTENTIAL.  
 FT ACT\_SITE 78 78 CHARGE RELAY SYSTEM (BY SIMILARITY).  
 FT ACT\_SITE 125 125 CHARGE RELAY SYSTEM (BY SIMILARITY).  
 FT ACT\_SITE 222 222 CHARGE RELAY SYSTEM (BY SIMILARITY).  
 FT DISULFID 26 145 INTERCHAIN (POTENTIAL).  
 FT DISULFID 63 79 BY SIMILARITY.  
 FT DISULFID 159 228 BY SIMILARITY.  
 FT DISULFID 192 210 BY SIMILARITY.  
 FT DISULFID 218 246 BY SIMILARITY.  
 FT CARBOHYD 85 85 N-LINKED (GLCNAC... ) (POTENTIAL).  
 FT VARIANT 60 60 M -> V (IN GAMMA-II).  
 FT VARIANT 126 126 /FTID=VAR\_012097.  
 FT VARIANT 132 132 I -> M (IN GAMMA-II).  
 FT VARIANT 132 132 S -> T (IN GAMMA-II).  
 FT VARIANT 204 204 /FTID=VAR\_012099.  
 FT VARIANT 288 288 L -> I (IN GAMMA-II).  
 FT VARIANT 288 288 /FTID=VAR\_012100.  
 FT VARIANT 288 288 L -> F (IN GAMMA-II).  
 FT CONFLICT 160 160 /FTID=VAR\_012101.  
 FT SEQUENCE 321 AA; 33827 MW; FFF7B06E3C4A962D CRC64;

Alignment Scores:  
 Pred. No.: 0.352 Length: 321  
 Score: 99.50 Matches: 49  
 Percent Similarity: 37.71% Conservative: 17  
 Best Local Similarity: 28.00% Mismatches: 50  
 Query Match: 3.70% Indels: 59  
 DB: 1 Gaps: 8

US-10-020-441-1 (1-1477) x TRYG\_HUMAN (1-321)  
 QY 52 TTCGGTTCATGCGATTCTTAACGACAGAGACA---ATGCTACAGSTTCATAGTC 108  
 DB 49 TTPTPTTPTGlnAlaSerLeuArgLeuArgMetHisValCysGlyGlySerLeu 68  
 QY 109 TCACGACGACGACTACTCAAGCTGCTATGTTTGTCTACCATTCGACGATTCG 168  
 DB 69 SerProGlnTrpValLeuThrAlaHisCysPheSerGlySerLeu-----Asn 85  
 QY 169 GTTAGAGATCGACGAAACACATGTTGCTCAATGTAATCGATTCATTCACAGACAG 228  
 DB 86 SerSerAspTyrGlnValHisLeu----- 93  
 QY 229 TTTCGCTGCGCGGCTTTGTGTTGTATGACGAGGGTTTCATAGATTATGTC 288  
 DB 94 -----GlyGlnLeuGlnIleThrLeuSer 101  
 QY 289 ATATACTGCTTATTATTCAGGTTTCATTCACATGAGAAATGGACCAACAGGC 348  
 DB 102 ProHis-----PheSerThrValArg-----GlnIleIle 111  
 QY 349 ATTCATACCAACCGTCTGAGTTAAGTGGACCAAGATATACGCTCTTGTATGTC 408  
 DB 112 LeuHisSerSerProSerGly-----GlnProGly----- 121  
 QY 409 GCACGACAGAGAGACCAATCGACACAGACACTGAGATTCGATATTCGATTGAATG 468  
 DB 122 -----ThSerGly-----AspIleAlaLeuValGln 130  
 QY 469 CTGGCTCAATATGCTCAACTTACAGAGTGAATCAGATGATCAGTCGCCACAGCATCG 528

DB 131 LeuSerValProValThrLeuSerSerArgIleLeuProValCysLeuProGlnAlaSer 150  
 QY 529 GATATCCGCCACCTGGAACGATGTTTCATTCGTTGTTGTTATGGA 573  
 DB 151 AspaSphneCysPProGlyIleArgCysTrpValThrGlyTrpGly 165  
 RESULT 32  
 MCT7\_MOUSE STANDARD; PRT; 273 AA.  
 AC 002844;  
 DT 01-JUL-1993 (Rel. 26, Created)  
 DT 01-JUL-1993 (Rel. 26, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE Mast cell protease 7 precursor (BC 3.4.21.59) (MMP-7) (Trypsinase).  
 GN MCP7.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OX NCBI\_TaxID=10090;  
 [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=DMB/2;  
 RX MEDLINE=93087489; PubMed=1454796;  
 RA McNeil H.P., Reynolds D.S., Schiller V., Childyal N., Gurley D.S.,  
 RA Austen K.F., Stevens R.L.;  
 RT "Isolation, characterization, and transcription of the gene encoding  
 RT mouse mast cell protease 7.";  
 RT Proc. Natl. Acad. Sci. U.S.A. 89:11174-11178(1992).  
 RN [2]  
 RP SEQUENCE FROM N.A. (TRUNCATED ISOFORM).  
 RC STRAIN=C57BL/6J;  
 RX MEDLINE=96162035; PubMed=8576265;  
 RA Hunt J.E., Stevens R.L., Austen K.F., Zhang J., Xia Z.,  
 RA Childyal N.;  
 RT "Natural disruption of the mouse mast cell protease 7 gene in the  
 RT C57BL/6 mouse.";  
 RL J. Biol. Chem. 271:2851-2855(1996).  
 CC -I- FUNCTION: TRYPTASE IS THE MAJOR NEUTRAL PROTEASE PRESENT IN MAST  
 CC CELLS, AND IS SECRETED UPON THE COUPLED ACTIVATION-DEGRANULATION  
 CC RESPONSE OF THIS CELL TYPE.  
 CC -I- CATALYTIC ACTIVITY: Preferential cleavage: Arg-|-, Lys-|-, but  
 CC with more restricted specificity than trypsin.  
 CC -I- ALTERNATIVE PRODUCTS: 2 isoforms, a long form (shown here) and a  
 CC truncated form; are produced by alternative splicing. The  
 CC alternative splicing event is due to a G to A point mutation at  
 CC the exon 2/intron 2 splice site and causes loss of protein  
 CC expression. The alternatively spliced transcript is only found in  
 CC C57BL/6 mouse.  
 CC -I- DEVELOPMENTAL STAGE: IS NOT EXPRESSED IN MATURE SEROSAL OR  
 CC MUCOSAL MAST CELLS AND IS EXPRESSED ONLY TRANSIENTLY AT AN EARLY  
 CC STAGE OF IN VITRO MAST CELL DIFFERENTIATION.  
 CC -I- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1. TRYPTASE SUBFAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (see <http://www.isb-sdb.ch/announce/>  
 CC or send an email to [license@isb-sdb.ch](mailto:license@isb-sdb.ch)).  
 CC -----  
 CC EMBL; L00654; AAA39993.1; -;  
 CC EMBL; L00653; AAA39992.1; -;  
 CC EMBL; U42405; AAA97874.1; -;  
 CC EMBL; U42406; AAA97875.1; -;  
 CC PIR; A47246; A47246.  
 CC HSSP; P20231; LAAO.  
 CC MEROPS; S01.026; -;  
 CC MGD; MGI:96943; Mpcp7.  
 DR InterPro; IPR001314; Chymotrypsin.  
 DR InterPro; IPR001254; Ser\_protease\_\_Try.  
 DR Pfam; PF00089; trypsin; 1.

DR PRINTS; PR00722; CHYMOTRYPSIN.  
 DR SMART; SMO0020; Tryp\_Spc; 1.  
 DR PROSITE; PS00240; TRYPSIN\_DOM; 1.  
 DR PROSITE; PS00134; TRYPSIN\_HIS; 1.  
 DR PROSITE; PS00135; TRYPSIN\_SER; 1.  
 DR Hydrolyase; Serine protease; Signal; Zymogen; Alternative splicing;  
 KM Glycoprotein.  
 FT SIGNAL 1 18 POTENTIAL.  
 FT PROPEP 19 28 ACTIVATION PEPTIDE (POTENTIAL).  
 FT CHAIN 29 273 MAST CELL PROTEASE 7 (POTENTIAL).  
 FT ACT\_SITE 72 72 CHARGE RELAY SYSTEM (BY SIMILARITY).  
 FT ACT\_SITE 119 119 CHARGE RELAY SYSTEM (BY SIMILARITY).  
 FT ACT\_SITE 222 222 CHARGE RELAY SYSTEM (BY SIMILARITY).  
 FT DISULFID 57 73 BY SIMILARITY.  
 FT DISULFID 153 228 BY SIMILARITY.  
 FT DISULFID 186 209 BY SIMILARITY.  
 FT DISULFID 218 246 BY SIMILARITY.  
 FT CARBOHYD 49 49 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 130 130 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT VASPLIC 44 46 VSL -> GCC (IN TRUNCATED ISOFORM).  
 FT VASPLIC 47 273 MISSING (IN TRUNCATED ISOFORM).  
 SQ SEQUENCE 273 AA; 30337 MW; 50BCB4765294205E CRC64;

Alignment Scores:  
 Pred. No.: 0.375 Length: 273  
 Score: 99.00 Matches: 37  
 Percent Similarity: 34.39% Conservative: 28  
 Best Local Similarity: 19.58% Mismatches: 56  
 Query Match: 3.69% Gaps: 68  
 DB: 1 Indels: 4

US-10-020-441-1 (1-1477) x MCT7\_MOUSE (1-273)

QY 40 CACCCACATGATTCGCCCTTCATCCGATCTTAACAGACAGAGAAC----- 87  
 DB 36 HisGlyAsnLysTyrProTyrGlnValSerLeuArgAlaAsnAspThrTyrTrpMetHis 55  
 QY 88 ATGTGTACAGSTTCACTCACTCAACAGAGACGATCAACAGCTGGTGTGTTGTC 147  
 DB 56 PheCysGlyGlySerLeuIleHisProGlnTyrPvalLeuThrAlaHisCysValGly 75  
 QY 148 TCACCATGGCCA-----GTGATTCGGGTAAAGATCGACTGAACACATTTGCT 198  
 DB 76 ProAspValAlaAspProAsnLysValArgValGlnLeuArgLysGlnTyrLeuTyrTyr 95  
 QY 199 CAATGATTCGATTCATTCACAGACAGCTGTTCGCTGCGTTCGCTTGTGTGTGTA 258  
 DB 96 HisAspHis----- 98  
 QY 259 TGCAGTGGGTTGCANAGATTATTAATGCTATATTAATGCTATTAATGCTTCAATTT 318  
 DB 99 -----LeuMetThrValSerGlnIleIleHisProAspPhe 111  
 QY 319 CTCGACGTGAGAAATGGCGACCAACAGGCATTCACCAACCGCTGGAATTAAGTG 378  
 DB 112 TyrIleValGlnAsp----- 116  
 QY 379 GCACACGATATACATGCCCTTGTATGTGCGACACAGACAGACCAATCGACAGACA 438  
 DB 116 ----- 116  
 QY 439 CTCAGTGGATTCGATTCGATTCGATTCGATTCGATTCGATTCGATTCGATTCGATTC 498  
 DB 117 -----GlyAlaAspIleAlaLeuLeuLysLeuThrAsnProValAsnIleSerAspTyr 134  
 QY 499 ATCAGAGTATGATCTGTCGACAGCCATCGATTCGATTCGATTCGATTCGATTCGATTC 558  
 DB 135 ValHisProValProLeuProProAlaSerGlnThrPheProSerClyThrLeuCysTyr 154  
 QY 559 ATGTGTGTTATGAGAGGATGATAC 585  
 DB 155 ValThrGlyTyrGlyAsnIleAspAsn 163

RESULT 33  
 ID MPN\_HUMAN STANDARD; PRT; 290 AA.  
 AC O990R3;  
 DT 16-OCT-2001 (Rel. 40, Created)  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE Marapsin precursor (EC 3.4.21.-).  
 MN MPN.  
 GN Homo sapiens (Human).  
 OS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RA Fortunate M., Dando P.M., Rawlings N.D., Barrett A.J.;  
 RT "Cloning, sequencing and expression of marapsin, a human serine  
 RL proteinase.";  
 CC Submitted (MAR-2001) to the EMBL/GenBank/DBJ databases.  
 CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL; AJ306593; CAC35467.1; -.  
 DR HSSP; P00734; IUVS.  
 DR MEROPS; S01.074; -.  
 DR InterPro; IPR001314; Chymotrypsin.  
 DR InterPro; IPR001254; Ser\_protease\_Try.  
 DR Pfam; PF00089; trypsin; 1.  
 DR PRINTS; PR00722; CHYMOTRYPSIN.  
 DR SMART; SMO0020; Tryp\_Spc; 1.  
 DR PROSITE; PS00240; TRYPSIN\_DOM; 1.  
 DR PROSITE; PS00134; TRYPSIN\_HIS; 1.  
 DR PROSITE; PS00135; TRYPSIN\_SER; 1.  
 KM Hydrolyase; Serine protease; Zymogen; Signal; Glycoprotein.  
 FT SIGNAL 1 22 POTENTIAL.  
 FT PROPEP 23 34 POTENTIAL.  
 FT CHAIN 35 290 MARAPIN.  
 FT DOMAIN 35 277 SERINE PROTEASE.  
 FT ACT\_SITE 75 75 CHARGE RELAY SYSTEM (BY SIMILARITY).  
 FT ACT\_SITE 124 124 CHARGE RELAY SYSTEM (BY SIMILARITY).  
 FT ACT\_SITE 229 229 CHARGE RELAY SYSTEM (BY SIMILARITY).  
 FT DISULFID 60 76 BY SIMILARITY.  
 FT DISULFID 158 235 BY SIMILARITY.  
 FT DISULFID 191 214 BY SIMILARITY.  
 FT DISULFID 225 253 BY SIMILARITY.  
 FT CARBOHYD 55 55 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 79 79 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 SQ SEQUENCE 290 AA; 31940 MW; 67BDC93EC70BFF7B CRC64;

Alignment Scores:  
 Pred. No.: 0.381 Length: 290  
 Score: 99.00 Matches: 49  
 Percent Similarity: 36.90% Conservative: 20  
 Best Local Similarity: 26.20% Mismatches: 48  
 Query Match: 3.69% Gaps: 70  
 DB: 1 Indels: 9

US-10-020-441-1 (1-1477) x MPN\_HUMAN (1-290)

QY 91 TGTACAGSTTCACTGATTCACAGACAGATCAACAGCTGTGTTGTTGTCGA 150  
 DB 60 CysGlyGlySerLeuIleAlaGlnTyrPvalLeuThrAlaHisCysPhe----- 77  
 QY 151 CCATTGCCAGTATTCGGGTAAAGATGCACTGAACACATTTGTTGCTCAATGTAATGCA 210  
 DB 78 -----Arg 78

```

QY 211 TTGATTTACACAGACGTTTGGCTGCGTGGCTGTTGTTGTTGATGACGGGTTT 270
    |||:::|
Db 79 AsnThrSerGluThr----- 83
QY 271 GCATGATTTTATGCTATATACCTTGCTTATTTACAGTTTCATTCTCACTGAGG 330
    |||:::|
Db 84 -----SerLeuTyrGlnValLeuLeu----- 90
QY 331 AATGGGACCAACAGGCGATCCATCCACCAACGCTGCGATTAGGTGGACCAAGATAC 390
    |||:::|
Db 91 ---GlyAlaAspGln-----LeuValGlnProGly--- 99
QY 391 ATGCCCTCTTGTATGTCGCGACGACAGAGAGA-----CCAAATCGACAG--- 435
    |||:::|
Db 100 ---ProHisAlaMetTyrAlaThrValAlaGlnValAlaGlnSerAsnProLeuTyrGlnGly 118
QY 436 ACACCTCAGTGGATTGCATATTGCAATTGTAATGCTGGCTCAAAATGGTCAACTTACAGAGT 495
    |||:::|
Db 119 ThrAlaSerSerAlaAspValAlaLeuValAlaGlnLeuValAlaProValProPheThrAsn 138
QY 496 GGAATCAGAGTATCATGCTGTCACAGCGCATGGATATCCCGCACCTGGAACTGGTGT 555
    |||:::|
Db 139 TyrLeuLeuProValCysLeuProAspProSerValIlePheGluThrGlyMetCys 158
QY 556 TTCATTGTTGTTATGGAAG---GATGTAACGACCGTGATCCGTACAGTAAGATGCT 612
    |||:::|
Db 159 TrpValIleThrGlyTyrPglYserProSerGluGluAspLeuLeuProGluProArg----- 176
QY 613 GGAATATTGAAGAAAGTGAAGT 633
    |||:::|
Db 177 ---IleLeuGlnIlyLeuAla 182

RESULT 34
VS89_CABEL
ID _YS89_CABEL STANDARD; PRT; 3178 AA.
AC Q09624; Q09625; Q0969D4;
DT 01-NOV-1995 (Rel. 32, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Hypothetical protein ZK945.9 in chromosome II.
GN ZK945.9/ZK945.10.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditioidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Bristol N2;
RA Wilkinson-Sprout J.;
RL Submitted (FEB-1995) to the EMBL/GenBank/DBJ databases.
RN [2]
RP REVISIONS.
RA Durin R.;
RL Submitted (DEC-2001) to the EMBL/GenBank/DBJ databases.
CC -1- SUBCELLULAR LOCATION: Integral membrane protein (potential).
CC -1- SIMILARITY: CONTAINS 1 GPS DOMAIN.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; Z48544; CAB70192.1; -.
DR EMBL; Z48582; CAB70192.1; JOINED.
DR EMBL; Z48582; CAB70201.1; -.
DR EMBL; Z48544; CAB70201.1; JOINED.
DR WormPeP; ZK945.9; CE25697.
DR InterPro; IPR002111; Cat_channel_TpL.
DR InterPro; IPR001024; Lipoxxygenase_LH2.

```

```

DR InterPro; IPR00636; M-channel_nlg.
DR InterPro; IPR000203; PKD_cys_rich.
DR Pfam; PF00520; Ion_trans; 1.
DR Pfam; PF01477; PLAT; 1.
DR SMART; SM00303; GPS; 1.
DR SMART; SM00308; LH2; 1.
DR Hypothetical protein; Transmembrane.
KM SMART; SM00308; LH2; 1.
FT DOMAIN 266 1196 SER/THR-RICH.
FT DOMAIN 1105 1241 GLY/SER-RICH.
FT DOMAIN 2071 2120 GPS.
FT TRANSMEM 13 30 POTENTIAL.
FT TRANSMEM 51 73 POTENTIAL.
FT TRANSMEM 2139 2161 POTENTIAL.
FT TRANSMEM 2348 2367 POTENTIAL.
FT TRANSMEM 2390 2412 POTENTIAL.
FT TRANSMEM 2451 2468 POTENTIAL.
FT TRANSMEM 2483 2505 POTENTIAL.
FT TRANSMEM 2567 2589 POTENTIAL.
FT TRANSMEM 2836 2858 POTENTIAL.
FT TRANSMEM 2939 2961 POTENTIAL.
FT TRANSMEM 2976 2998 POTENTIAL.
FT TRANSMEM 3038 3060 POTENTIAL.
SQ SEQUENCE 3178 AA; 344726 MW; F8239436D03666CD CRC64;

Alignment Scores:
Pred. No.: 0.709 Length: 3178
Score: 99.00 Matches: 69
Percent Similarity: 39.11% Conservative: 37
Best Local Similarity: 25.46% Mismatches: 107
Query Match: 3.71% Indels: 58
DB: Gaps: 11

US-10-020-441-1 (1-1477) x VS89_CABEL (1-3178)
QY 1399 TGTGACCATGTGATGACACCGGATGACTGACCTGAAGGATGGAGAGAGCTCC 1340
    |||:::|
Db 267 CysAsp-----TyrThrIleGluSerThrSerThrThrThrProThrThr 283
QY 1339 ACCACGTGACCTGGAGCGGATGACGCGAAATTCGACCTGCTTCACACATATAGG 1280
    |||:::|
Db 284 ThrThrValThrSerThr-----ValThrSerThrThrThrValProThrSerThr 301
QY 1279 ATGGCATTGGTCGCGATGTCGACATTCATTATAGTCGACCTGCGACCTGTGTGCAACAA 1220
    |||:::|
Db 302 ThrValThrThrAlaMetSerThrSerThrSerThrProSerThrSerThrThrIleGlu 321
QY 1219 ACACCCACACACAGTACAGTACACAAACGTAACGTAACATGAGAAAGAAACAAACAC 1160
    |||:::|
Db 322 SerThrSerThrThr-Phe---ThrSerThrAlaSerThrSerThrSerThrSerThr 340
QY 1159 AACGGACAAACAAACAAACACACAGTGAATTCGAAATATCATCTGCACATATCCAC 1100
    |||:::|
Db 340 ThrGlnGlnSerSerSerThrIleThrSerSerProSerSerThr---ThrLeuSerThr 359
QY 1099 CCTTTACACTCCACACACACA-----ATATCCACATAC-----A 1064
    |||:::|
Db 359 SerIleProThrThrThrThrThrProGluIleThr-SerThrLeuSerSerLeuProAspA 379
QY 1063 ATCATGTGACCTATATTTCACGATCAAGTC-----ACTCTCATCTCACCCTCTCC 1010
    |||:::|
Db 379 snAlaIleCysSerTyrLeuAspGluThrThrThrSerThrThrPheThrThrMetL 399
QY 1009 AAACATATCTGAACACACATGACCCCTCACTTCA----- 975
    |||:::|
Db 399 eutThrSerThrThrThrGluGluProSerThrSerThrThrThrGluValThrSerT 419
QY 974 -----TGTACACACACACA 962
    |||:::|
Db 419 hSerSerThrValThrThrThrGluProThrThrThrLeuThrThrSerThrAlaSerT 439
QY 961 CTTTCATCGTTTGACATGATTAATTTCATGAGCTTCTCTCCCAAGTGGTCTCATCATC 902
    |||:::|

```



Db 439 hrSerThrGluProSerThrSerThrValThrThrSerProSer-----ThrSerP 457  
 QY 901 CCATCATCATCATTCATCCATCCATCTCACTCACTCACTCACTCACTCACTCACT 842  
 |||||  
 Db 457 rovalThrSerThrValThrSerSerSerSerSerSerThr-----ValThrT 474  
 QY 841 CACCTACCAACACAAATGACCGAGATGAAATCACTCTGCTCACTGAACTGCACAAAGTC 782  
 |||||  
 Db 474 hrPro-----ThrSerThrGluSerThrSerThrS 484  
 QY 781 ATCCAAACACACTGTGTATTATCATCAAGCTCTGACCAATCAACACCGATAGTAGAC 722  
 |||||  
 Db 484 erProSerSerThrValThrThrSerThrThrAlaProSerThrSerThrThrGlyProS 504  
 QY 721 AGACAGACACACAAACAGATACACAGACA 693  
 |||||  
 Db 504 erSerSerSerSerThrProSerSerThr 513  
 RESULT 35  
 MCT2\_RAT STANDARD; PRT; 247 AA.  
 ID MCT2\_RAT  
 AC P00770:  
 DT 21-JUL-1986 (Rel. 01, Created)  
 DT 01-JAN-1988 (Rel. 06, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Mast cell protease II precursor (BC 3.4.21.-) (RMCP-II) (RMCP-2)  
 OS (Group-specific protease).  
 OS Rattus norvegicus (Rat).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
 ON NCB1\_TaxID=10116;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=87165980; PubMed=3549719;  
 RA Benfey P.N., Yin F.H., Leder P.;  
 RT "Cloning of the mast cell protease, RMCP II. Evidence for  
 RT cell-specific expression and a multi-gene family.";  
 RL J. Biol. Chem. 262:5377-5384(1987).  
 RN [2]  
 RP SEQUENCE OF 21-244.  
 RX MEDLINE=78124137; PubMed=629933;  
 RA Woodbury R.G., Katunuma N., Kobayashi K., Titani K., Neurath H.;  
 RT "Covalent structure of a group-specific protease from rat small  
 RT intestine. Appendix: crystallographic data for a group specific  
 RT protease from rat intestine.";  
 RL Biochemistry 17:811-819(1978).  
 RN [3]  
 RP X-RAY CRYSTALLOGRAPHY (1.9 ANGSTROMS).  
 RX MEDLINE=89166471; PubMed=3233198;  
 RA Remington S.J., Woodbury R.G., Reynolds R.A., Mathews B.W.,  
 RA Neurath H.;  
 RT "The structure of rat mast cell protease II at 1.9-A resolution.";  
 RL Biochemistry 27:8097-8105(1988).  
 CC -1- FUNCTION: THIS ENZYME, ISOLATED FROM SMALL INTESTINE, SPECIFICALLY  
 CC INACTIVATES THE APO FORMS OF A CERTAIN GROUP OF INTRACELLULAR  
 CC PYRIDOXAL PHOSPHATE-REQUIRING ENZYMES. IT HAS CHYMOTRYPSIN-LIKE  
 CC SPECIFICITY TOWARDS SMALL SUBSTRATES.  
 CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1. GRANZYME SUBFAMILY.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See http://www.isb-sib.ch/announce/  
 CC or send an email to license@isb-sib.ch).  
 CC  
 CC EMBL: J02712; AAA66284.1; -  
 CC PIR: A00957; PRRG.  
 CC PIR: A29546; A29548.  
 CC PDB: 3RP2; 09-JAN-89.  
 CC MEROPS: S01.141; -  
 CC InterPro: IPR001314; Chymotrypsin.

DR InterPro: IPR001254; Ser\_protease\_Try.  
 DR Pfam: PF00089; trypsin\_1.  
 DR PRINTS: PR00722; CHYMOTRYPSIN.  
 DR SMART: SM00020; Tryp\_SPC; 1.  
 DR PROSITE: PS50240; TRYPSIN\_DOM; 1.  
 DR PROSITE: PS00134; TRYPSIN\_HIS; 1.  
 DR PROSITE: PS00135; TRYPSIN\_SER; 1.  
 DR Hydrolase; Serine protease; 3D-structure; Zymogen; signal;  
 KW Multigene family.  
 FT SIGNAL 1 18  
 FT PROPEP 19 20 ACTIVATION PEPTIDE.  
 FT CHAIN 21 247 MAST CELL PROTEASE II.  
 FT ACT\_SITE 65 65 CHARGE RELAY SYSTEM.  
 FT ACT\_SITE 109 109 CHARGE RELAY SYSTEM.  
 FT ACT\_SITE 202 202 CHARGE RELAY SYSTEM.  
 FT DISULFID 50 66  
 FT DISULFID 143 208  
 FT DISULFID 174 187  
 FT CONFLICT 238 238 W -> T (IN REF. 2).  
 FT STRAND 22 22  
 FT STRAND 25 26  
 FT TURN 29 30  
 FT TURN 33 34  
 FT STRAND 35 41  
 FT TURN 43 44  
 FT STRAND 49 54  
 FT STRAND 57 62  
 FT HELIX 64 66  
 FT STRAND 70 75  
 FT STRAND 79 79  
 FT TURN 80 81  
 FT TURN 85 86  
 FT TURN 88 97  
 FT TURN 99 100  
 FT STRAND 111 115  
 FT STRAND 122 122  
 FT TURN 123 124  
 FT STRAND 125 125  
 FT TURN 133 134  
 FT TURN 139 140  
 FT STRAND 142 147  
 FT STRAND 150 152  
 FT TURN 153 154  
 FT STRAND 155 157  
 FT STRAND 160 160  
 FT STRAND 162 169  
 FT HELIX 171 173  
 FT TURN 174 178  
 FT TURN 182 184  
 FT STRAND 185 188  
 FT TURN 191 192  
 FT STRAND 196 196  
 FT TURN 199 203  
 FT STRAND 205 208  
 FT TURN 209 210  
 FT STRAND 211 218  
 FT TURN 221 222  
 FT STRAND 227 231  
 FT HELIX 232 243  
 SQ SEQUENCE 247 AA; 27101 MW; 051988042A97A7E CRC64;  
 Alignment Scores:  
 Pred. No.: 0.406 Length: 247  
 Score: 98.50 Matches: 45  
 Percent Similarity: 33.83% Conservative: 23  
 Best Local Similarity: 22.39% Mismatches: 64  
 Query Match: 3.67% Indels: 69  
 DB: 1 Gaps: 7  
 US-10-020-441-1 (1-1477) x MCT2\_RAT (1-247)  
 QY 13 TTGATACGTAGTGGAACCTGTGCAACACCGCACTGAAATCCGCTTCATCGCATTTCTTA 72  
 |||||  
 |||||

Db 21 IleIleGlyValGluSerIleProHisSerArg-----ProTyrMetAlaHisLeu 38  
 QY 73 -----ACGACAGAG-----AGAACATGTGTACAGSTTCACTAGTCTCAACGAGA 117  
 Db 39 AspIleValThrGluLeuGlyLeuArgValIleCysGlyGlyPheLeuIleSerArgIn 58  
 QY 118 GCAGTACTACAGCGTGTGTCATTTGTCACCATTCGCGAGTATTCGGGTAAAGAGT 177  
 Db 59 PheValLeuThrAlaAlaHisCysGlyArgGluIleThrValIleLeuGly----- 76  
 QY 178 CGACTGAAACATGTGTCTCATATGATTCATTCATTCACAGACAGTGTTCGGTGT 237  
 Db 77 -----AlaHisAspValArgLys----- 82  
 QY 238 GCGTCGCTTGTGTGTGTATGATGAGTGGTTCATAGATTTTATGCTATATCTTG 297  
 Db 82 ----- 82  
 QY 298 GTCTTATTTACAGTTTTCATTTCTCACACTGAGGAATGGGCAACACAGCATTCATCAC 357  
 Db 83 -----ArgGluSerThrGlnGlnLysIleVal 92  
 QY 358 CAACCGCTCGAGTTAAGGTGCACACAGATACATGCCCTCTTGTATGTCGACAGACAG 417  
 Db 93 GlnLysGlnIleIleHisGluSerTyrAsnSerValProAsnLeu----- 107  
 QY 418 AGGAGACCAATGCACACAGACTCAGTGTGATTCGATTCATTTGTAATGCTGCTCAA 477  
 Db 108 -----HisAspIleMetLeuIleLysLeuGlnLys 117  
 QY 478 ATGTCTCACTTACAGAGTGAATCAGAGTATGATCTGCAACAGCATTCGATATCCG 537  
 Db 118 LysValGluLeuThrProAlaValAsnValValProLeuProSerProSerAspPheIle 137  
 QY 538 CCACCTGCACTGGTGTTCATTTGTTGTTATGAGAGGAGATACACACCGTATCCG 597  
 Db 138 HisProGlyAlaMetCysTrpAlaAlaGlyTrpGlyLys---ThrGlyValArgAspPro 156  
 QY 598 TCA 600  
 Db 157 Thr 157  
 RESULT 36  
 CYAL DROME STANDARD: PRT: 2248 AA.  
 AC P32870:  
 DT 01-OCT-1993 (Rel. 27, Created)  
 DT 01-OCT-1993 (Rel. 27, Last sequence update)  
 DT 30-MAY-2000 (Rel. 39, Last annotation update)  
 DE Ca2+/calmodulin-responsive adenylyl cyclase (EC 4.6.1.1) (ATP  
 DE pyrophosphate-lyase) (Rutabaga protein).  
 GN RUT.  
 OS Drosophila melanogaster (Fruit fly).  
 OC Eukaryota; Metazoa; Arthropoda; Mandibulata; Pancrustacea; Hexapoda;  
 OC Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;  
 OC Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.  
 OX NCBI\_TaxId=7227;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Canton-S; TISSUE=Head;  
 RX MEDLINE=9215464; PubMed=173965;  
 RA Levin L.R., Han P.-L., Wang P.M., Feinstein P.G., Davis R.L.,  
 RA Reed R.R.;  
 RT "The Drosophila learning and memory gene rutabaga encodes a  
 RT Ca2+/calmodulin-responsive adenylyl cyclase.";  
 RL Cell 68:479-489 (1992).  
 CC -1- FUNCTION: THIS A MEMBRANE-BOUND, CALMODULIN-SENSITIVE ADENYLYL  
 CC CYCLASE. INACTIVATION OF THIS CYCLASE LEADS TO A LEARNING AND  
 CC MEMORY DEFECT.  
 CC -1- CATALYTIC ACTIVITY: ATP = 3',5'-cyclic AMP + diphosphate.  
 CC -1- ENZYME REGULATION: ACTIVATED BY CA(2+)/CALMODULIN AND G PROTEIN.  
 CC -1- SUBCELLULAR LOCATION: Integral membrane protein.  
 CC -1- TISSUE SPECIFICITY: MUSHROOM BODIES OF THE FLY BRAIN.

CC -1- DOMAIN: COMPOSED OF TWO HOMOLOGOUS DOMAINS.  
 CC -1- SIMILARITY: BELONGS TO ADENYLYL CYCLASE CLASS-4/GUANYLYL CYCLASE  
 CC FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb.ch/announce/>  
 CC or send an email to [license@isb.ch](mailto:license@isb.ch)).  
 CC -----  
 CC EMBL: M81887; AAA28844.1; -.  
 DR PIR: D42088; D42088.  
 DR HSSP: P19754; 1AMK.  
 DR FlyBase: FBgn0003301; rut.  
 DR InterPro: IPR001054; G\_cyclase.  
 DR Pfam: PF00211; guanylate\_cyc; 2.  
 DR SMART: SM00044; CYCC; 2.  
 DR PROSITE: PS00452; GUANYLATE\_CYCLASES\_1; 2.  
 DR PROSITE: PS0125; GUANYLATE\_CYCLASES\_2; 2.  
 KW lysase; cAMP synthesis; Transmembrane; Glycoprotein; Repeat.  
 FT DOMAIN 1 41  
 FT TRANSMEM 42 60 POTENTIAL.  
 FT TRANSMEM 65 84 POTENTIAL.  
 FT TRANSMEM 101 115 POTENTIAL.  
 FT TRANSMEM 122 142 POTENTIAL.  
 FT TRANSMEM 152 174 POTENTIAL.  
 FT TRANSMEM 186 206 POTENTIAL.  
 FT DOMAIN 207 705  
 FT TRANSMEM 706 726 POTENTIAL.  
 FT TRANSMEM 730 750 POTENTIAL.  
 FT TRANSMEM 770 791 POTENTIAL.  
 FT DOMAIN 792 813  
 FT TRANSMEM 814 834 POTENTIAL.  
 FT TRANSMEM 842 867 POTENTIAL.  
 FT TRANSMEM 868 888 POTENTIAL.  
 FT TRANSMEM 889 2248  
 FT DOMAIN 318 334 CATALYTIC (POTENTIAL).  
 FT DOMAIN 1013 1029 CATALYTIC (POTENTIAL).  
 FT DOMAIN 515 530 GLY-RICH.  
 FT DOMAIN 569 602 GLY-RICH.  
 FT DOMAIN 1278 1297 GLN-RICH.  
 FT DOMAIN 1767 1810 GLY/SER-RICH.  
 FT DOMAIN 2025 2040 ASP/GLU-RICH (ACIDIC).  
 FT DOMAIN 2241 GLN-RICH.  
 FT CARBOHYD 800 800 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 807 807 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT MUTAGEN 1026 1026 G->R; ABOLISHES CATALYTIC ACTIVITY.  
 SQ SEQUENCE 2248 AA; 248899 MW; E459C718B01868 CRC64;  
 Alignment Scores:  
 Pred. No.: 0.721 Length: 2248  
 Score: 98.50 Matches: 49  
 Percent Similarity: 36.13% Conservative: 20  
 Best Local Similarity: 25.65% Mismatches: 65  
 Query Match: 3.69% Indels: 57  
 DB: 1 Gaps: 9  
 US-10-020-441-1 (1-1477) x CYAL DROME (1-2248)  
 QY 1120 ATCACTGCGACAAATATCCACCTTTACACTACCCACCAACAATAGTCCATCAATC 1061  
 Db 1425 LeuAsnGlnHisGlyHisHisProHisHisLeuHisSerAsn----- 1439  
 QY 1060 ACATGTGCATTAATTCAGCATCAAGTCACTCTCCATTCACCACTCTCCAAACATA 1001  
 Db 1440 ---LeuAsnLeuAsnGlnSerGlnHisProProSerPheThrSerLeuGlyTyrGlyGln 1458  
 QY 1000 CTCAAACACATGACCCCTTACTTCACTATGACACACACACTTCATCGTTTGATGATA 941  
 Db 1459 CysArgGluSerGluProLeuHisAla----- 1468

QY 940 TTTCATGAGTCTTCTCTCCAGTGGTCTCAATCATCCATCATCATTCATCC 881  
 DB 1469 -----SerSerValAlaProValAlaLysIleMetProMetGlnHisAlaProLys 1485  
 QY 880 CATCCAACTTACTGACCTACCTACCTACCTACCTACCTACCTACCTACCTAC 821  
 DB 1486 Tyr-----GluProProArgTyrThrSerProHis----- 1495  
 QY 820 GAGATGAAATCATCTCTACTGAACTGACAAAGTCAAAACAGATGTTATTA 761  
 DB 1486 -----ThimetleuserGlnGlnHisGlnGlnGlnHis----- 1508  
 QY 760 TCAAGGCTTCACCAATACACACACCGATGACAGACAGACAGACAGACAGATA 701  
 DB 1509 -----GlnHisGlnGlnPro--Gln--SerGlnSerGlnSerAlaGlnAsp 1523  
 QY 700 CACAGACACACAGAAATACATGATCATCTGACATGACATGACATGATGATGATTC 641  
 DB 1523 GlnGlnHisProAlaGlnAspProHisPro-----LeuG 1535  
 QY 640 ACCAACAATCACTTCTCTCAT-----ATTCAC-----CCATTCTTACGG 599  
 DB 1555 LnaGlnTyrLametyLysSerGlnGlnProGlnLeuProProLysProValLeuArgT 1555  
 QY 598 ACGGATCAAGGTGTTATCATCCCTTCCA 570  
 DB 1555 hTyrMetLysProLeuProLysLeuPro 1564  
 RESULT 37  
 YH17 YEAST  
 ID YH17 YEAST STANDARD; PRT; 153 AA.  
 AC P38898;  
 DT 01-FEB-1995 (Rel. 31, Created)  
 DT 01-FEB-1995 (Rel. 31, Last sequence update)  
 DT 01-OCT-1996 (Rel. 34, Last annotation update)  
 DE Hypothetical 17.1 KDa protein in PURS 3' region.  
 GN YHR217C.  
 OS Saccharomyces cerevisiae (Baker's Yeast).  
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
 OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.  
 OX NCBI\_TaxID=4932;  
 RN (1)  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=5288C / AB972;  
 RX MEDLINE=94378003; PubMed=8091229;  
 RA Johnson M., Andrews S., Brinkman R., Cooper J., Ding H., Dover J.,  
 Du Z., Favello A., Fulton L., Gateung S., Geisel C., Kirsten J.,  
 Kucaba T., Hillier L., Jier M., Johnston L., Langston Y.,  
 Latreille P., Louis E.J., Macri C., Mardis B., Meneses S., Mouser L.,  
 Nhan M., Rifkin L., Riles L., St Peter H., Trevasakis B., Vaughan K.,  
 Vignati D., Wilcox L., Wohldman P., Waterston R., Wilson R.,  
 Vaudin M.,  
 RT "Complete nucleotide sequence of Saccharomyces cerevisiae chromosome  
 VIII."  
 RL Science 265:2077-2082(1994).  
 CC -1- SIMILARITY: TO YEAST YNL338M.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See http://www.isb-sib.ch/announce/  
 CC or send an email to license@isb-sib.ch).  
 CC -----  
 CC EMBL; U00029; AAB69740.1; -  
 DR PIR; S48998; S48998.  
 DR SGD; S0001260; YHR217C.  
 KW Hypothetical protein.  
 SQ SEQUENCE 153 AA; 17080 MW; B632686D08E74049 CRC64;  
 Alignment Scores: 0.399 Length: 153  
 Pred. No.:

Score: 98.00 Matches: 41  
 Percent Similarity: 45.77% Conservative: 24  
 Best Local Similarity: 28.87% Mismatches: 40  
 Query Match: 3.68% Indels: 37  
 DB: 1 Gaps: 8  
 US-10-020-441-1 (1-1477) x YH17 YEAST (1-153)  
 QY 1141 ACAGAGTGAATTTGAAATATTCATGACCAATATTCACCCCTTACACTACCCAC 1082  
 DB 10 SerSerMetGlnTyrSerAspIleTyrIleHisThrProHisProHisProHis 29  
 QY 1081 AACCAATAGTCACATATACATCATGATGACATATATTCAGAGATCACTTCCATC 1022  
 DB 30 ProHisThrPro--ThrHisThrHisProHisThrProHisThrProHisThrProHis 48  
 QY 1021 TCACCACTCTGCAACACATATCTCAAC--ACATGACCCCTTACTTCACTACACAC 965  
 DB 48 SerHisThrProHisProHisThrHisThrHisProHisThrHisThrHisThrHis 68  
 QY 964 ACACCTTCATCGTTTGATGATATTCATGATGATGATGATGATGATGATGATGATG 905  
 DB 68 His-----ThrHisLeuSerHis-LeuSerLeuAsn- 78  
 QY 904 ATCCCATCATCATCATTCATCCATCCATCCATCATCTCACTCACTCACTCACTCA 845  
 DB 79 LeuProSerHisTyr--ProHisSerProLeu--ValThrLeuProHisSerThrIle 96  
 QY 844 CTT-----CACTCAGCAAC-----ACA 827  
 DB 97 ProLeuProThrThrIleHisLeuSerThrTyrTyrTyrHisProProProIleIleThr 116  
 QY 826 ATGACC-----GAGATGAAATATCATCTCTGCTCACTGATGATGATGATGAT 788  
 DB 117 ValThrLeuGlnLeuProIleSerAsnSerThrThrIleThrLeuLeuProTyrHis 136  
 RESULT 38  
 MUC3 HUMAN  
 ID MUC3 HUMAN STANDARD; PRT; 213 AA.  
 AC Q02505; Q02506;  
 DT 01-JUN-1994 (Rel. 29, Created)  
 DT 01-JUN-1994 (Rel. 29, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE Mucin 3A (Intestinal mucin 3A) (Fragments).  
 GN MUC3A OR MUC3.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN (1)  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Small Intestine;  
 RX MEDLINE=90365738; PubMed=2393399;  
 RA Gum J.R., Jr., Hicks J.W., Swallow D.M., Lagace R.L., Byrd J.C.,  
 Lampert D.T.A., Siddiki B., Kim Y.S.;  
 RT "Molecular cloning of cDNAs derived from a novel human intestinal  
 mucin gene."  
 RL Biochem. Biophys. Res. Commun. 171:407-415(1990).  
 CC -1- FUNCTION: MAJOR GLYCOPROTEIN COMPONENT OF A VARIETY OF MUCUS GELS.  
 CC THOUGHT TO PROVIDE A PROTECTIVE, LUBRICATING BARRIER AGAINST  
 CC PARTICLES AND INFECTIOUS AGENTS AT MUCOSAL SURFACES.  
 CC -1- SUBCELLULAR LOCATION: Secreted.  
 CC -1- TISSUE SPECIFICITY: BROAD SPECIFICITY; SMALL INTESTINE, COLON AND  
 CC COLONIC TUMORS.  
 CC -----  
 CC -1- PTM: HIGHLY O-GLYCOSYLATED AND PROBABLY ALSO N-GLYCOSYLATED.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See http://www.isb-sib.ch/announce/  
 CC or send an email to license@isb-sib.ch).



Score:	98.00	Matches:	36
Percent Similarity:	35.71%	Conservative:	22
Best Local Similarity:	21.43%	Mismatches:	44
Query Match:	3.65%	Indels:	66
DB:	1	Gaps:	4

```
Db      46  user1acysgiylglypheleuValarg1uaspheValleuthrAlaAlaHisCysLe 66
QY      144  TTGCTCACCATTTGCCAGT-----ATTGGGTAAAGATCGACTGAA 185
Db      66  uGlyserSerIleasnValThrleuGlyAlaHisasnIleGlnMetArgGluArgThrGI 86
QY      186  ACACATTGTGTGCTCAATGATTCATTTTCACAGACAGTGTTCGCGTGCCTGCGT 245
Db      86  nGlnleuLe-----ThrValLeu----- 92
QY      246  TTGTGTTGTGTATGCAGTGGGTTGCATAGATTTTAATGTCATATACTTGGTCTTATT 305
Db      92  ----- 92
QY      306  TCAGGTTTCATTTTCACACTGAGGAATGGCGACCAACAAGGATCCATCACCACCGTC 365
Db      93  -----ArgAlaIleArgHis----- 97
QY      366  TGGAGTTAAGTGGCACACGATACATGCCCTTGTGTATGTCGGCACAGAGGAGACC 425
Db      98  -----ProaspIytrAsnPro----- 102
QY      426  AATCGCACAGACACTCGAGTGGATTCGATTTTGCAATTGTGAATGCTGGCTCAAAATGCTCA 485
Db      103  -----GlnasnIleArgAsn--AspIleMetLeuLeuGlnLeuArgArgAlaAr 119
QY      486  CTTACAGAGTGAATCGAGTGTATCGCTCGCCACAGCCATCGGATATCCCGCACCTGG 545
Db      119  gArgSerGIyservalysProvalAlaLeuProGlnAlaSerLysLysLeuGlnProGI 139
QY      546  AACTGGTGTTCATTTGTTGTTATGGAAGGATGATTAACGACCGT 591
Db      139  yAspIeuCysThrvalAlaGlyTrpGlyArgValserGlnserArg 154
```

Search completed: April 9, 2003, 13:31:35  
Job time : 63.5 secs

GenCore version 5.1.4.P5.4578  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: April 9, 2003, 13:09:27 ; Search time 6.81967 Seconds  
(without alignments)  
225.546 Million cell updates/sec

Title: US-10-020-441-3

Perfect score: 16

Sequence: 1 VGGGRDDNDPDRSKN 16

Scoring table: OLIGO

Gapop 60.0 , Gapext 60.0

Searched: 283224 seqs, 96134422 residues

Word size : 0

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Listing first 1000 summaries

Database :

1: pir1:\*  
2: pir2:\*  
3: pir3:\*  
4: pir4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	16	100.0	264	2	A28942
2	7	43.8	686	2	E87450
3	6	37.5	112	2	D69543
4	6	37.5	122	2	G69532
5	6	37.5	160	2	AG0689
6	6	37.5	243	2	S74031
7	6	37.5	244	2	A98263
8	6	37.5	292	2	AF3032
9	6	37.5	303	2	F98253
10	6	37.5	335	2	AB2193
11	6	37.5	412	2	A70839
12	6	37.5	450	2	S73419
13	6	37.5	509	2	S74935
14	6	37.5	1592	2	S63208
15	6	37.5	45	2	S24713
16	5	31.2	63	2	G90800
17	5	31.2	75	2	D81239
18	5	31.2	76	2	H72412
19	5	31.2	80	2	A91036
20	5	31.2	80	2	C85880
21	5	31.2	80	2	F65011
22	5	31.2	97	2	JQ0041
23	5	31.2	101	2	S64056
24	5	31.2	107	2	T39293
25	5	31.2	109	2	T29627
26	5	31.2	110	2	C95956
27	5	31.2	110	2	PC4215
28	5	31.2	121	2	A27888
29	5	31.2	121	2	A13574

30	5	31.2	125	2	B69287	conserved hypotet
31	5	31.2	127	2	E87434	hypothetical prote
32	5	31.2	128	2	S76955	hypothetical prote
33	5	31.2	128	2	A75199	hypothetical prote
34	5	31.2	133	2	A96706	unknown protein, 6
35	5	31.2	134	2	PC4214	phosphatidylethano
36	5	31.2	144	2	AG3272	probable lactoylgl
37	5	31.2	141	2	AF0812	probable acetyltra
38	5	31.2	151	2	A97008	transcription regu
39	5	31.2	152	2	PC4216	phosphatidylethano
40	5	31.2	152	2	A44947	lek antigen precu
41	5	31.2	152	2	A95023	hypothetical prote
42	5	31.2	158	2	AB1442	probable repressor
43	5	31.2	160	1	G70591	glucanokinase [imp
44	5	31.2	160	2	AG2191	acetylactate synth
45	5	31.2	163	1	YCEC3H	acetylactate synth
46	5	31.2	163	2	B90639	acetylactate synth
47	5	31.2	163	2	B85490	conserved hypotet
48	5	31.2	165	2	F69819	s18 protein - frui
49	5	31.2	166	2	S13219	hypothetical prote
50	5	31.2	167	2	C97856	hypothetical prote
51	5	31.2	167	2	S17400	hypothetical prote
52	5	31.2	167	2	S49246	hypothetical prote
53	5	31.2	171	2	G86480	hypothetical prote
54	5	31.2	172	2	C97894	DNA binding protei
55	5	31.2	174	2	I40416	hypothetical prote
56	5	31.2	178	2	F87408	hypothetical prote
57	5	31.2	184	2	S06695	hypothetical prote
58	5	31.2	187	2	A97605	hypothetical prote
59	5	31.2	187	2	AC2827	acetyltransferase
60	5	31.2	191	2	B84207	probable acetyltra
61	5	31.2	192	2	S43563	R01H0.5 protein -
62	5	31.2	192	2	E72855	AcOrf-45 protein -
63	5	31.2	200	2	B69859	hypothetical prote
64	5	31.2	206	2	S68597	flavin reductase (
65	5	31.2	207	2	B69085	conserved hypotet
66	5	31.2	209	2	A48388	glutathione S-tran
67	5	31.2	209	2	D88577	protein R01H0.5 (
68	5	31.2	211	2	F86674	thymidylate kinase
69	5	31.2	215	2	T17207	hypothetical prote
70	5	31.2	222	2	T26500	peptidylprolyl iso
71	5	31.2	223	2	G95351	viral type IV secr
72	5	31.2	224	2	H72346	basal-body rod mod
73	5	31.2	227	2	H96566	hypothetical prote
74	5	31.2	229	2	T34277	hypothetical prote
75	5	31.2	235	2	AE3344	integral membrane
76	5	31.2	242	2	S71757	MADS box protein D
77	5	31.2	243	2	A96744	hypothetical prote
78	5	31.2	244	2	A87436	arsenical resistan
79	5	31.2	250	2	S40405	MADS box protein o
80	5	31.2	253	2	A38233	triose-phosphate i
81	5	31.2	258	2	B95998	conserved hypotet
82	5	31.2	265	2	T02138	hypothetical prote
83	5	31.2	266	2	F69176	conserved hypotet
84	5	31.2	271	1	Z8BPC2	gene 18 protein -
85	5	31.2	275	2	JW0104	azarene carbazole
86	5	31.2	282	2	T43119	hypothetical prote
87	5	31.2	282	2	T26397	hypothetical prote
88	5	31.2	284	2	S26451	hypothetical prote
89	5	31.2	284	2	S26437	hypothetical prote
90	5	31.2	287	2	T01919	probable Ap2 domai
91	5	31.2	289	2	S60968	probable membrane
92	5	31.2	289	2	JC4577	transcription elon
93	5	31.2	299	2	S53395	hypothetical prote
94	5	31.2	302	2	A31921	collagen dpy-13 pr
95	5	31.2	306	2	H89791	hypothetical prote
96	5	31.2	307	2	F97639	hypothetical prote
97	5	31.2	307	2	AC2862	conserved hypotet
98	5	31.2	309	2	E82137	transcription regu
99	5	31.2	309	2	A95988	conserved hypotet
100	5	31.2	311	2	S13808	protein-tyrosine k
101	5	31.2	311	2	AF0448	probable HlyD fami
102	5	31.2	311	2	S23533	hypothetical prote

103	5	31.2	317	2	S44151	cathepsin L (EC 3.
104	5	31.2	320	2	C86289	hypothetical prote
105	5	31.2	324	2	F69504	proliferating-cell
106	5	31.2	324	2	S40432	IG4-specific metal
107	5	31.2	326	2	C86832	phosphate acetyltr
108	5	31.2	330	2	F75410	conserved hypotet
109	5	31.2	335	2	T8161	heat shock protein
110	5	31.2	336	2	AG0105	sugar transport sy
111	5	31.2	336	2	F86319	Hypothetical prote
112	5	31.2	340	2	C75004	hypothetical prote
113	5	31.2	346	2	S13807	protein-tyrosine k
114	5	31.2	349	2	D64134	peptide transport
115	5	31.2	365	2	H96647	hypothetical prote
116	5	31.2	365	2	JQ1146	C-5 steroid desatur
117	5	31.2	366	2	D90290	1-carnitine deshydr
118	5	31.2	366	2	A86392	hypothetical prote
119	5	31.2	368	2	S17980	pgal protein - Asp
120	5	31.2	369	2	A56492	protein kinase ERK
121	5	31.2	370	2	S31831	IG4-specific metal
122	5	31.2	370	2	S40431	IG4-specific metal
123	5	31.2	371	1	E64821	ylit protein precu
124	5	31.2	371	2	E90743	probable dehydroge
125	5	31.2	371	2	A85594	probable dehydroge
126	5	31.2	372	2	D87022	conserved hypotet
127	5	31.2	375	2	B96743	unknown protein [i
128	5	31.2	377	2	A99710	conserved hypotet
129	5	31.2	380	2	A86712	hypothetical prote
130	5	31.2	383	2	S24156	polylacturonase
131	5	31.2	385	2	T19201	hypothetical prote
132	5	31.2	387	2	E11404	cone arrestin - Af
133	5	31.2	387	2	E15423	arrestin-C - human
134	5	31.2	388	2	S38943	arrestin - human
135	5	31.2	389	2	S68172	cone arrestin - hu
136	5	31.2	389	2	S68175	cone arrestin - bu
137	5	31.2	392	2	T45032	hypothetical prote
138	5	31.2	393	2	T07653	acyl-l-acyl-carrier
139	5	31.2	393	2	S44202	acyl-l-acyl-carrier
140	5	31.2	394	2	T51494	stearoyl-acyl carr
141	5	31.2	394	2	B82931	hypothetical membr
142	5	31.2	396	1	OHCSAD	acyl-l-acyl-carrier
143	5	31.2	396	1	A39173	acyl-l-acyl-carrier
144	5	31.2	396	2	B39170	acyl-l-acyl-carrier
145	5	31.2	396	2	T14172	acyl-l-acyl-carrier
146	5	31.2	396	2	T14268	acyl-l-acyl-carrier
147	5	31.2	398	2	C95278	hypothetical prote
148	5	31.2	398	2	AC1163	peptidoglycan lyti
149	5	31.2	399	1	OHSPAD	acyl-l-acyl-carrier
150	5	31.2	399	2	S24995	acyl-l-acyl-carrier
151	5	31.2	400	2	B83028	proteinhase subunit
152	5	31.2	401	2	T32345	hypothetical prote
153	5	31.2	401	2	A11387	peptidoglycan lyti
154	5	31.2	403	2	S42532	hypothetical prote
155	5	31.2	406	2	D75272	nifR3 protein - De
156	5	31.2	406	2	A75183	hypothetical prote
157	5	31.2	407	2	H84920	probable Tub faml1
158	5	31.2	407	2	T22554	hypothetical prote
159	5	31.2	409	1	A47527	transcription fact
160	5	31.2	409	2	H71030	hypothetical prote
161	5	31.2	411	2	T07806	acyl-l-acyl-carrier
162	5	31.2	414	2	G84385	hypothetical prote
163	5	31.2	414	2	T43851	translation initia
164	5	31.2	418	2	A12852	poly(A) polymerase
165	5	31.2	420	2	T01172	GI/S transition co
166	5	31.2	420	2	E15088	L-SP precursor - J
167	5	31.2	424	1	XNBYCP	choline-phosphate
168	5	31.2	427	1	A55496	apexin precursor -
169	5	31.2	434	2	F83643	conserved hypotet
170	5	31.2	435	2	G97629	polA polymerase (p
171	5	31.2	436	2	F69101	phosphomannomutase
172	5	31.2	437	2	S41766	heterogeneous nucl
173	5	31.2	438	2	H70844	probable aminotran
174	5	31.2	439	2	E60916	HNF-3/forhead hom
175	5	31.2	439	2	T49189	kinesin heavy chai
176	5	31.2	441	2	G64516	hypothetical prote
177	5	31.2	445	2	D83819	hypothetical prote
178	5	31.2	447	2	A95980	probable two-compo
179	5	31.2	448	1	VH1HBC	nucleocapsid prote
180	5	31.2	448	1	VH1HN1	nucleocapsid prote
181	5	31.2	448	1	A60003	nucleocapsid prote
182	5	31.2	448	2	JQ1173	nucleocapsid prote
183	5	31.2	454	2	G70476	conserved hypotet
184	5	31.2	456	2	B83609	probable aminotran
185	5	31.2	457	2	T29741	hypothetical prote
186	5	31.2	458	2	A95893	probable aminotran
187	5	31.2	459	2	S17742	protochlorophyllid
188	5	31.2	460	2	E72341	cysteine-tRNA liga
189	5	31.2	461	2	F40605	hmc 3'-region hypo
190	5	31.2	462	2	S34620	envelope protein -
191	5	31.2	467	1	T07565	probable protochlo
192	5	31.2	467	2	B87637	antioxidantferase,
193	5	31.2	467	2	AG0771	probable two-compo
194	5	31.2	467	2	AE3142	hypothetical prote
195	5	31.2	469	2	B87094	probable molecular
196	5	31.2	470	1	A25685	hypothetical prote
197	5	31.2	470	2	T39796	negative sexual co
198	5	31.2	474	2	A71533	probable phospholi
199	5	31.2	475	2	AC3449	chitoooligosacchari
200	5	31.2	476	2	S64953	hypothetical prote
201	5	31.2	478	2	D64230	NADH oxidase (nox)
202	5	31.2	479	2	S73770	NADH oxidase nox -
203	5	31.2	479	2	F83014	hypothetical prote
204	5	31.2	486	2	AC2474	hypothetical prote
205	5	31.2	489	1	T01204	sulfate adenylyltr
206	5	31.2	493	1	ACR1B1	nucleotide acetylch
207	5	31.2	495	2	G70593	probable cationic
208	5	31.2	496	2	D97728	thermostable carbo
209	5	31.2	500	2	T26786	hypothetical prote
210	5	31.2	501	2	AG3356	glutamate-tRNA lig
211	5	31.2	508	1	NDBER3	exonuclease (EC 3.
212	5	31.2	515	2	T39448	hypothetical prote
213	5	31.2	516	2	G87425	alpha-L-arabinofur
214	5	31.2	517	2	S36712	FN26 protein - ye
215	5	31.2	520	2	G98145	probable aminotran
216	5	31.2	525	2	T48824	hypothetical prote
217	5	31.2	535	2	A69537	cytochrome oxidase
218	5	31.2	536	2	AH3227	genec protein [imp
219	5	31.2	537	2	A84670	probable nucleotid
220	5	31.2	544	2	H85064	4-comarnate-CoA li
221	5	31.2	553	2	C75318	hypothetical prote
222	5	31.2	555	2	T29377	hypothetical prote
223	5	31.2	555	2	B98235	conserved hypotet
224	5	31.2	555	2	AG3050	hypothetical prote
225	5	31.2	556	2	G86319	F25116.5 protein -
226	5	31.2	558	2	C90695	probable transport
227	5	31.2	558	2	G85545	probable membrane
228	5	31.2	558	2	E64778	probable potassium
229	5	31.2	563	2	AC0376	probable potassium
230	5	31.2	564	2	D96017	probable peptidase
231	5	31.2	569	1	VHXPLJ	major structural n
232	5	31.2	569	2	A36187	interleukin-1 rece
233	5	31.2	569	2	F75381	probable two-compo
234	5	31.2	569	2	T35476	probable regulator
235	5	31.2	573	2	D90202	methionyl-tRNA syn
236	5	31.2	573	2	C84645	hypothetical prote
237	5	31.2	576	2	A32604	interleukin-1 rece
238	5	31.2	584	2	T01173	GI/S transition co
239	5	31.2	587	2	H95370	probable transport
240	5	31.2	587	2	J01650	SGLI protein - hum
241	5	31.2	588	2	B87521	peptidase M1 faml1
242	5	31.2	591	2	B95963	probable acetolact
243	5	31.2	592	2	T47821	methylenetetrahydr
244	5	31.2	593	1	K1ZPMN	protein kinase cdr
245	5	31.2	595	1	S20825	protein-tyrosine-p
246	5	31.2	595	1	A44390	protein-tyrosine-p
247	5	31.2	595	2	E69470	amino-acid ABC tra
248	5	31.2	595	2	T16140	hypothetical prote



249	5	31.2	601	2	C89451	protein T04G9.6 [l	322	5	31.2	889	2	JC6015	chitin synthase (E
250	5	31.2	604	2	C97541	chloride channel,	323	5	31.2	904	2	C70559	probable po1a prot
251	5	31.2	604	2	AE2760	hypothetical prote	324	5	31.2	908	2	AH0055	probable cation-tr
252	5	31.2	607	2	B84837	hypothetical prote	325	5	31.2	926	2	E96749	probable lipoxigen
253	5	31.2	608	2	AC2796	hypothetical prote	326	5	31.2	926	2	T16184	hypothetical prote
254	5	31.2	608	2	C97575	glutamine-fructose	327	5	31.2	928	2	T38948	hypothetical cole
255	5	31.2	609	2	C87106	glutamine penicilli	328	5	31.2	950	2	T22592	hypothetical prote
256	5	31.2	609	2	AB0500	glutamine-fructose	329	5	31.2	960	2	T24221	hypothetical prote
257	5	31.2	610	1	S20444	leukorrhene-A4 hyd	330	5	31.2	976	2	S35457	SNF2 protein homol
258	5	31.2	611	1	S65947	leukorrhene-A4 hyd	331	5	31.2	990	2	D83706	lanthibiotic mersac
259	5	31.2	611	1	JN0066	leukorrhene-A4 hyd	332	5	31.2	997	2	T43523	cutr17 protein - fi
260	5	31.2	611	1	JC4237	leukorrhene-A4 hyd	333	5	31.2	1000	2	S44898	ZK1236.3 protein -
261	5	31.2	612	2	JT0592	hypothetical prote	334	5	31.2	1007	2	C84668	probable receptor-
262	5	31.2	612	2	G69797	conserved hypotet	335	5	31.2	1015	2	C84918	probable ATP-depen
263	5	31.2	625	2	B96710	hypothetical prote	336	5	31.2	1027	2	A56533	chromatin remodel
264	5	31.2	627	1	AD1818	hypothetical prote	337	5	31.2	1034	2	A95262	probable formate d
265	5	31.2	632	1	GVN85X	N-acetyluramoyl-L	338	5	31.2	1077	2	A96533	probable zinc meta
266	5	31.2	635	2	AF1003	surface glycoprote	339	5	31.2	1091	2	T30256	calcium channel al
267	5	31.2	637	2	S75772	probable ABC trans	340	5	31.2	1099	2	H83210	probable phosphol
268	5	31.2	639	2	T16648	hypothetical prote	341	5	31.2	1106	1	A39299	DNA-directed DNA p
269	5	31.2	640	2	AF2178	hypothetical prote	342	5	31.2	1114	2	T30299	dynein heavy chain
270	5	31.2	643	2	AD2471	penicillin-binding	343	5	31.2	1120	2	S67208	hypothetical prote
271	5	31.2	643	2	AF2471	penicillin-binding	344	5	31.2	1166	1	S06142	protein-tyrosine k
272	5	31.2	643	2	G89008	protein T27C4.2 [l	345	5	31.2	1169	2	T30207	dynein heavy chain
273	5	31.2	651	2	S82299	cell division prot	346	5	31.2	1204	2	T19918	hypothetical prote
274	5	31.2	653	2	A82334	DNA mismatch repai	347	5	31.2	1209	2	T52523	hypothetical prote
275	5	31.2	657	2	C89815	methionyl-tRNA syn	348	5	31.2	1217	2	T51140	disease resistance
276	5	31.2	658	2	T40107	hypothetical 57.9	349	5	31.2	1217	2	T51141	disease resistance
277	5	31.2	659	2	F70175	rep helicase, sing	350	5	31.2	1250	2	T30301	dynein heavy chain
278	5	31.2	659	2	E64594	penicillin-binding	351	5	31.2	1259	2	A43425	Bravo/Nr-CAM cell
279	5	31.2	660	2	F71917	penicillin-binding	352	5	31.2	1263	2	T00649	hypothetical prote
280	5	31.2	667	2	T28847	hypothetical prote	353	5	31.2	1268	1	A39640	neural cell adhesi
281	5	31.2	671	2	T23015	hypothetical prote	354	5	31.2	1274	2	T10729	transferrin-like p
282	5	31.2	675	2	G71640	NADH2 dehydrogenas	355	5	31.2	1337	2	T13948	atypical protein k
283	5	31.2	676	2	G84663	hypothetical prote	356	5	31.2	1337	2	T16718	hypothetical prote
284	5	31.2	677	2	J50178	protein kinase YKR	357	5	31.2	1357	2	S57052	hypothetical prote
285	5	31.2	683	2	T21810	hypothetical prote	358	5	31.2	1361	2	A29959	DNA-directed RNA p
286	5	31.2	686	2	T21808	hypothetical prote	359	5	31.2	1364	2	T14900	condensin XCAP-D2
287	5	31.2	691	2	B87383	cell cycle histidi	360	5	31.2	1387	2	JC5502	G-protein signalin
288	5	31.2	698	2	D81832	transferrin-bindin	361	5	31.2	1438	2	S59792	probable membrane
289	5	31.2	710	2	T49516	Actu related protei	362	5	31.2	1467	2	T23950	hypothetical prote
290	5	31.2	713	2	T40051	hypothetical prote	363	5	31.2	1468	2	S11515	formin - mouse
291	5	31.2	717	2	A39203	coma protein - Str	364	5	31.2	1478	2	C82689	helicase, ATP depe
292	5	31.2	717	2	C97877	transport ATP-bind	365	5	31.2	1532	2	A26039	hypothetical prote
293	5	31.2	717	2	P95004	hypothetical prote	366	5	31.2	1545	2	P98262	hypothetical prote
294	5	31.2	718	2	T29448	hypothetical prote	367	5	31.2	1561	2	S61314	glutamate synthase
295	5	31.2	719	2	T52510	hypothetical prote	368	5	31.2	1581	2	AD3022	hypothetical prote
296	5	31.2	723	2	C64058	outer membrane pro	369	5	31.2	1582	2	T15308	hypothetical prote
297	5	31.2	725	2	A57148	outer membrane pro	370	5	31.2	1635	2	T32452	protein vit-6 [imp
298	5	31.2	727	2	P95422	catalase (EC 1.11.	371	5	31.2	1651	2	P88750	vitellogenin vit-6
299	5	31.2	727	2	AB2353	hypothetical prote	372	5	31.2	1651	2	B43081	etarch synthase DU
300	5	31.2	744	2	T11035	hypothetical prote	373	5	31.2	1674	2	T01265	protein kinase - f
301	5	31.2	750	2	A97501	topoisomerase IV s	374	5	31.2	1706	2	T39305	hypothetical prote
302	5	31.2	750	2	AE2719	topoisomerase IV s	375	5	31.2	1771	2	S76851	IgA-specific metal
303	5	31.2	755	2	A44315	carliage oligomer	376	5	31.2	1773	2	A81937	major mezozote su
304	5	31.2	764	2	AB1695	p60-relaxed protei	377	5	31.2	1785	2	A45546	IgA-specific metal
305	5	31.2	768	2	S43567	ROH10.7 protein (	378	5	31.2	1815	2	C81169	IgA-specific metal
306	5	31.2	774	2	H83057	penicillin-binding	379	5	31.2	1819	2	B86410	protein F3M18.14 l
307	5	31.2	775	2	T41961	hypothetical prote	380	5	31.2	2023	2	T13154	polycomb protein e
308	5	31.2	777	2	A82303	penicillin-binding	381	5	31.2	2055	2	S29529	genome polypotein
309	5	31.2	781	2	D64138	penicillin-binding	382	5	31.2	2195	2	T30197	alpha tectorin - m
310	5	31.2	788	2	T64045	recombination prot	383	5	31.2	2157	2	A13009	peptide synthetase
311	5	31.2	796	2	E86654	hypothetical prote	384	5	31.2	2175	1	S03170	homeotic protein c
312	5	31.2	797	2	T06219	probable cyet nema	385	5	31.2	2201	1	GNN7A9	genome polypotein
313	5	31.2	802	2	H75420	hypothetical prote	386	5	31.2	2485	1	H71621	serine/threonine-s
314	5	31.2	803	1	RGBYW6	regulatory protein	387	5	31.2	2554	1	TVF7L1	kinase-related pro
315	5	31.2	812	2	P88577	protein R01H0.7 [	388	5	31.2	2566	2	E98274	hypothetical prote
316	5	31.2	814	2	S65083	5-methyltetrahydro	389	5	31.2	2657	2	T18497	hypothetical prote
317	5	31.2	820	2	C75284	phenylalanyl-tRNA	390	5	31.2	3229	2	S27852	probable cell surf
318	5	31.2	826	1	A31822	villin - chicken	391	5	31.2	6831	2	A88852	protein unc-22 [lm
319	5	31.2	827	1	A31642	villin [validated]	392	5	31.2	6839	2	S57242	twichin [seimilari
320	5	31.2	833	2	S45041	genome polypotein	393	5	31.2	7160	2	T27935	hypothetical prote
321	5	31.2	885	2	B70393	hypothetical prote	394	5	31.2	8243	2	T31307	type I fatty acid

395	4	25.0	15	2	PHI342	Ig heavy chain DJ	468	4	25.0	70	2	AC2866	hypothetical prote
396	4	25.0	15	2	G49255	T-cell receptor be	469	4	25.0	71	1	W2VZ46	8.6k HindIII-C pro
397	4	25.0	18	2	S14661	photosystem I prot	470	4	25.0	71	1	T17598	hypothetical prote
398	4	25.0	20	2	A38566	Gmp-binding protei	471	4	25.0	71	2	S75276	hypothetical prote
399	4	25.0	25	2	S78341	hypothetical prote	472	4	25.0	71	2	G72497	hypothetical prote
400	4	25.0	29	2	T34643	hypothetical prote	473	4	25.0	71	2	AG2853	hypothetical prote
401	4	25.0	30	2	S68618	histone H2B - sea	474	4	25.0	71	2	AE3450	hypothetical prote
402	4	25.0	33	2	E46027	glycine transporte	475	4	25.0	72	1	A45393	late l2 mu core pr
403	4	25.0	33	2	F87516	hypothetical prote	476	4	25.0	72	2	T12130	potassium channel
404	4	25.0	36	2	H48110	RNA recognition mo	477	4	25.0	72	2	T03391	potassium channel
405	4	25.0	38	2	T11950	photosystem II pro	478	4	25.0	72	2	T22081	hypothetical prote
406	4	25.0	38	2	S78330	photosystem II pro	479	4	25.0	72	2	E97714	hypothetical prote
407	4	25.0	38	2	S73157	photosystem II pro	480	4	25.0	72	2	H97665	hypothetical prote
408	4	25.0	38	2	T06924	photosystem II pro	481	4	25.0	72	2	AD2890	hypothetical prote
409	4	25.0	39	2	S10315	photosystem II pro	482	4	25.0	73	2	AG60136	keratin, scale (cl
410	4	25.0	41	2	C72429	hypothetical prote	483	4	25.0	73	2	C89016	protein B0213.4 [i
411	4	25.0	41	2	S19566	ornatin A2 - leech	484	4	25.0	73	2	AG3473	lactoylglycethione
412	4	25.0	41	2	S19621	ornatin A3 - leech	485	4	25.0	73	2	AB2468	secretory protein
413	4	25.0	46	2	A60495	T-cell receptor be	486	4	25.0	74	2	E90799	hypothetical prote
414	4	25.0	47	2	A48505	lectin, sialic aci	487	4	25.0	74	2	A41711	coleoptericin - be
415	4	25.0	49	2	S66455	taurochenodeoxycho	488	4	25.0	75	2	A89016	hypothetical prote
416	4	25.0	49	2	S09049	potassium channel	489	4	25.0	75	2	E70870	hypothetical prote
417	4	25.0	50	2	S68843	sucrose 1F-fructos	490	4	25.0	76	2	T32666	hypothetical prote
418	4	25.0	51	2	B90274	transposase, ISC104	491	4	25.0	76	2	T43204	hypothetical prote
419	4	25.0	51	2	S78668	hypothetical prote	492	4	25.0	77	2	D86401	protein T22C5.7 [i
420	4	25.0	54	2	S59909	14-3-3 protein iso	493	4	25.0	77	2	A82970	hypothetical prote
421	4	25.0	56	2	B39406	dihydroliipoamide S	494	4	25.0	77	2	AC2101	lipoprotein (impor
422	4	25.0	56	2	A25507	proteinase inhibit	495	4	25.0	77	2	AG2343	hypothetical prote
423	4	25.0	57	2	S09048	potassium channel	496	4	25.0	78	2	H82796	hypothetical prote
424	4	25.0	57	2	S09042	potassium channel	497	4	25.0	79	2	C95254	D-alanyl carrier p
425	4	25.0	57	2	S09043	potassium channel	498	4	25.0	79	2	A43732	B6 protein - mdg
426	4	25.0	57	2	S09044	potassium channel	499	4	25.0	79	2	B82523	hypothetical prote
427	4	25.0	57	2	S09045	potassium channel	500	4	25.0	79	2	T46591	hypothetical prote
428	4	25.0	57	2	S09047	potassium channel	501	4	25.0	79	2	A99519	D-alanyl carrier p
429	4	25.0	57	2	D70255	hypothetical prote	502	4	25.0	80	2	B84285	hypothetical prote
430	4	25.0	58	2	C97808	hypothetical prote	503	4	25.0	80	2	E81899	hypothetical prote
431	4	25.0	60	2	S09765	hypothetical prote	504	4	25.0	80	2	AC0807	conserved hypotet
432	4	25.0	61	2	F87582	hypothetical prote	505	4	25.0	83	2	E69062	ferredoxin - Metha
433	4	25.0	62	1	KRSH16	major high-(glycin	506	4	25.0	83	2	T35979	hypothetical prote
434	4	25.0	62	2	T18774	hypothetical prote	507	4	25.0	83	2	T03201	hypothetical prote
435	4	25.0	62	2	H84255	hypothetical prote	508	4	25.0	83	2	AP2147	hypothetical prote
436	4	25.0	62	2	T51850	RING-H2 finger pro	509	4	25.0	84	2	E72065	ribosomal protein
437	4	25.0	62	2	G75004	hypothetical prote	510	4	25.0	84	2	E86558	L27 ribosomal prot
438	4	25.0	62	2	B82790	hypothetical prote	511	4	25.0	84	2	E44989	dmp protein - phag
439	4	25.0	63	2	A58511	bromocontryphan pr	512	4	25.0	84	2	D81721	hypothetical prote
440	4	25.0	63	2	S09643	probable cytochrom	513	4	25.0	85	2	T08595	cysteine proteinas
441	4	25.0	63	2	H86251	protein P25C20.5 [	514	4	25.0	85	2	T32664	hypothetical prote
442	4	25.0	63	2	F36856	B1SR protein - vari	515	4	25.0	85	2	H70229	hypothetical prote
443	4	25.0	63	2	G72173	D4R protein - Vari	516	4	25.0	86	2	G90652	inhibitor of Rho [
444	4	25.0	64	2	F86693	hypothetical prote	517	4	25.0	86	2	S01842	nlfw protein - K1e
445	4	25.0	64	2	A86888	preprotein translo	518	4	25.0	86	2	G85503	hypothetical prote
446	4	25.0	64	2	B82209	hypothetical prote	519	4	25.0	86	2	D98327	cell division topo
447	4	25.0	64	2	S10640	gnaA protein - Str	520	4	25.0	86	2	A12955	yeao protein - Esc
448	4	25.0	65	2	B81165	50S ribosomal prot	521	4	25.0	86	2	E64743	hypothetical prote
449	4	25.0	65	2	D89016	protein B0213.3 [i	522	4	25.0	87	2	T30424	hypothetical prote
450	4	25.0	65	2	B84354	hypothetical prote	523	4	25.0	87	2	S77415	hypothetical prote
451	4	25.0	65	2	E72375	preprotein translo	524	4	25.0	87	2	PC6014	hypothetical 87 pr
452	4	25.0	65	2	T31044	hypothetical prote	525	4	25.0	87	2	T26867	hypothetical prote
453	4	25.0	65	2	B71158	probable nusa prot	526	4	25.0	88	2	AG1356	hypothetical prote
454	4	25.0	66	2	T14884	hypothetical prote	527	4	25.0	88	2	AH1726	hypothetical prote
455	4	25.0	66	2	S09855	hypothetical prote	528	4	25.0	88	2	F81390	hypothetical prote
456	4	25.0	66	2	AH1908	hypothetical prote	529	4	25.0	89	2	T119805	hypothetical prote
457	4	25.0	67	2	D70929	probable ferredoxi	530	4	25.0	89	2	A59100	hypothetical prote
458	4	25.0	67	2	T18012	hypothetical prote	531	4	25.0	89	2	E70072	hypothetical prote
459	4	25.0	67	2	S14854	hypothetical prote	532	4	25.0	89	2	S55780	Hox C6 protein - e
460	4	25.0	68	2	PQ0817	methionine adenosy	533	4	25.0	89	2	T03334	gene e33 protein -
461	4	25.0	68	2	G95229	hypothetical prote	534	4	25.0	89	2	H82562	hypothetical prote
462	4	25.0	68	2	B98094	hypothetical prote	535	4	25.0	90	2	AB1459	B. subtilis yaaZ p
463	4	25.0	69	2	S48150	chlecystokinin 69	536	4	25.0	90	2	AG1095	B. subtilis yaaZ p
464	4	25.0	69	2	B89016	protein B0213.5 [i	537	4	25.0	90	2	D38179	hypothetical prote
465	4	25.0	70	2	S48170	chlecystokinin 70	538	4	25.0	90	2	AD2730	hypothetical prote
466	4	25.0	70	2	G45186	homeotic protein G	539	4	25.0	91	2	T05920	probable cysteine
467	4	25.0	70	2	A83982	hypothetical prote	540	4	25.0	91	2	T12081	late embryogenesis

541	4	25.0	91	2	T24120	hypothetical prote	614	4	25.0	104	2	E69405	hypothetical prote
542	4	25.0	91	2	T24562	hypothetical prote	615	4	25.0	104	2	T40938	very hypothetical
543	4	25.0	91	2	F81996	hypothetical prote	616	4	25.0	104	2	H82793	transposase OrfA x
544	4	25.0	91	2	T32667	hypothetical prote	617	4	25.0	104	2	H83530	hypothetical prote
545	4	25.0	91	2	S61907	protein YCR054c ho	618	4	25.0	104	2	H96497	probable auxin-ind
546	4	25.0	92	2	H47116	orf1-3' of tfgx -	619	4	25.0	104	2	A69037	conserved hypotnet
547	4	25.0	92	2	A05208	hypothetical prote	620	4	25.0	104	2	C72637	hypothetical prote
548	4	25.0	93	2	H97802	hypothetical prote	621	4	25.0	105	2	A45641	acidic ribosomal p
549	4	25.0	93	2	A99311	hypothetical prote	622	4	25.0	105	2	T38409	very hypothetical
550	4	25.0	94	1	A40499	defensin alpha-1 p	623	4	25.0	105	2	S09766	hypothetical prote
551	4	25.0	94	1	B40499	defensin alpha-3 p	624	4	25.0	105	2	B55519	mcRB protein - Str
552	4	25.0	94	1	F64364	conserved hypotnet	625	4	25.0	105	2	B86350	protein YIK7_27 [i
553	4	25.0	94	2	A99431	conserved hypotnet	626	4	25.0	105	2	F86774	hypothetical prote
554	4	25.0	94	2	T43127	conserved hypotnet	627	4	25.0	105	2	D83242	hypothetical prote
555	4	25.0	94	2	T46349	hypothetical prote	628	4	25.0	105	2	S52603	hypothetical prote
556	4	25.0	94	2	T30720	hypothetical prote	629	4	25.0	106	2	S25631	acidic ribosomal p
557	4	25.0	95	2	H72271	ferredoxin - Therm	630	4	25.0	106	2	S31435	acidic ribosomal p
558	4	25.0	95	2	D97296	stage V sporulatio	631	4	25.0	106	2	T24117	hypothetical prote
559	4	25.0	95	2	B71571	hypothetical prote	632	4	25.0	106	2	C90975	hypothetical prote
560	4	25.0	95	2	AG2439	hypothetical prote	633	4	25.0	106	2	T50636	hypothetical prote
561	4	25.0	96	2	B26074	cysteine proteinas	634	4	25.0	107	1	RSUT38	acidic ribosomal p
562	4	25.0	96	2	S17608	Ig heavy chain V r	635	4	25.0	107	2	T02234	arginine decarboxy
563	4	25.0	96	2	S08639	homeotic protein z	636	4	25.0	107	2	S59918	acidic ribosomal p
564	4	25.0	96	2	S15152	hypothetical prote	637	4	25.0	107	2	S59919	acidic ribosomal p
565	4	25.0	96	2	F83555	hypothetical prote	638	4	25.0	107	2	S59920	acidic ribosomal p
566	4	25.0	96	2	C81940	hypothetical prote	639	4	25.0	107	2	S40106	acidic ribosomal p
567	4	25.0	96	2	E70811	hypothetical prote	640	4	25.0	107	2	B70650	probable emrE prot
568	4	25.0	96	2	B83656	transcription plet	641	4	25.0	107	2	D83780	hypothetical prote
569	4	25.0	97	2	T52052	ubiquitin-conjugat	642	4	25.0	107	2	H83596	hypothetical prote
570	4	25.0	97	2	A42575	Ig kappa chain V r	643	4	25.0	108	1	L5H0DL	Ig lambda chain V-
571	4	25.0	97	2	A57983	outer membrane por	644	4	25.0	108	1	S01844	fibroin - silkworm
572	4	25.0	97	2	B90725	hypothetical prote	645	4	25.0	108	1	C69151	conserved hypotnet
573	4	25.0	97	2	CG1025	hypothetical prote	646	4	25.0	108	2	A49893	probable iron-regu
574	4	25.0	97	2	S26982	hypothetical prote	647	4	25.0	109	2	B49831	nitrogenase (EC 1.
575	4	25.0	97	2	C22848	hypothetical prote	648	4	25.0	109	2	S68171	Ig lambda chain V
576	4	25.0	97	2	C85576	hypothetical ORF-2	649	4	25.0	110	2	A61061	actindidin (EC 3.4
577	4	25.0	97	2	D97482	hypothetical prote	650	4	25.0	110	2	A69389	SSU ribosomal prot
578	4	25.0	97	2	F64809	hypothetical prote	651	4	25.0	110	2	AG1149	conserved hypotnet
579	4	25.0	98	2	D70069	ybbG protein - Bac	652	4	25.0	110	2	AH1508	conserved hypotnet
580	4	25.0	98	2	AH3327	bicyclomycin resis	653	4	25.0	110	2	G70021	hypothetical prote
581	4	25.0	99	2	S44658	ZK53_2 protein -	654	4	25.0	110	2	T38245	hypothetical prote
582	4	25.0	99	2	S35055	serum antitoxic pr	655	4	25.0	110	2	H91156	hypothetical prote
583	4	25.0	99	2	B75019	ssu ribosomal prot	656	4	25.0	111	2	S72737	adenine phosphorib
584	4	25.0	99	2	H96016	hypothetical prote	657	4	25.0	111	2	T28502	hypothetical prote
585	4	25.0	99	2	H95400	probable non-heme	658	4	25.0	111	2	F72158	IDL protein - vari
586	4	25.0	100	2	H70654	probable urea prot	659	4	25.0	111	2	H42511	G3L 12.8K protein
587	4	25.0	100	2	S33338	prolamine P2 - red	660	4	25.0	111	2	S33079	G3L protein - vari
588	4	25.0	100	2	T29453	hypothetical prote	661	4	25.0	111	2	E84325	hypothetical prote
589	4	25.0	100	2	E43550	hypothetical prote	662	4	25.0	111	2	T34565	hypothetical prote
590	4	25.0	101	2	E37476	E4 orf 3' of fiber	663	4	25.0	111	2	H82597	hypothetical prote
591	4	25.0	101	2	S33929	hypothetical prote	664	4	25.0	111	2	E86002	hypothetical prote
592	4	25.0	101	2	G97567	hypothetical prote	665	4	25.0	112	2	G70747	probable nitrogen
593	4	25.0	101	2	AR2788	hypothetical prote	666	4	25.0	112	2	C87493	nitrogen regulator
594	4	25.0	102	1	T09380	ribosomal protein	667	4	25.0	112	2	D87415	nitrogen regulator
595	4	25.0	102	1	EB4414	30S ribosomal prot	668	4	25.0	112	2	T35668	nitrogen regulator
596	4	25.0	102	2	S06612	chorion protein sl	669	4	25.0	112	2	H72682	hypothetical prote
597	4	25.0	102	2	AF1211	hypothetical prote	670	4	25.0	112	2	T35026	probable membrane
598	4	25.0	102	2	AF1567	hypothetical prote	671	4	25.0	113	2	S13494	major oleosin chai
599	4	25.0	102	2	AC2490	hypothetical prote	672	4	25.0	113	2	T36694	hypothetical prote
600	4	25.0	102	2	AB2492	hypothetical prote	673	4	25.0	113	2	A55205	xls 5'-region hypo
601	4	25.0	102	2	AB3553	hypothetical prote	674	4	25.0	113	2	T36789	hypothetical prote
602	4	25.0	102	2	D71042	hypothetical prote	675	4	25.0	113	2	E72648	hypothetical prote
603	4	25.0	102	2	EB9808	hypothetical prote	676	4	25.0	113	2	E72701	hypothetical prote
604	4	25.0	102	2	S67204	hypothetical prote	677	4	25.0	114	1	UC1155	ribonuclease P (EC
605	4	25.0	103	1	K5RB	Ig kappa-B5 chain	678	4	25.0	114	1	KVRBAH	Ig kappa chain V r
606	4	25.0	103	2	S52599	probable membrane	679	4	25.0	114	2	H87313	nitrogen regulator
607	4	25.0	103	2	A32167	homeotic protein H	680	4	25.0	114	2	C86305	hypothetical prote
608	4	25.0	103	2	H90234	hypothetical prote	681	4	25.0	114	2	C85676	unknown protein en
609	4	25.0	103	2	A42523	A5R protein - vac	682	4	25.0	114	2	B97485	hypothetical prote
610	4	25.0	103	2	U01791	Salifer protein -	683	4	25.0	115	2	T16069	hypothetical prote
611	4	25.0	104	1	K5RBV	Ig kappa chain C r	684	4	25.0	115	2	D81024	hypothetical prote
612	4	25.0	104	1	A41902	arsenical resistan	685	4	25.0	115	2	T43081	transfer complex p
613	4	25.0	104	2	S56385	hypothetical 11.8K	686	4	25.0	115	2	T21833	hypothetical prote

687	4	25.0	116	2	S14211	NADH2 dehydrogenas	760	4	25.0	125	2	AE1047	conserved hypotnet
688	4	25.0	116	2	G97689	glrk protein (Aa00	761	4	25.0	125	2	B70981	hypothetical prote
689	4	25.0	116	2	AD2915	nitrogen regulator	762	4	25.0	125	2	F87685	hypothetical prote
690	4	25.0	116	2	S06614	chorion protein s1	763	4	25.0	125	2	T16247	hypothetical prote
691	4	25.0	116	2	AB2553	hypothetical prote	764	4	25.0	126	2	T34337	hypothetical prote
692	4	25.0	116	2	AS2478	hypothetical prote	765	4	25.0	126	2	F72729	hypothetical prote
693	4	25.0	116	2	AE3493	hypothetical prote	766	4	25.0	126	2	A69250	hypothetical prote
694	4	25.0	116	2	G89050	protein T05C3.1 [i	767	4	25.0	126	2	S58331	probable membrane
695	4	25.0	116	2	D96530	probable ribosomal	768	4	25.0	127	1	LZPY	lysosome (BC 3.2.1
696	4	25.0	117	2	S28161	nerve growth facto	769	4	25.0	127	2	PC2268	cytochrome P450 pr
697	4	25.0	117	2	AB3201	conserved hypotnet	770	4	25.0	127	2	S62375	aspartic proteinas
698	4	25.0	118	1	MAWLB2	E4 protein - bovin	771	4	25.0	127	2	S00751	placental protein
699	4	25.0	118	2	B87341	response regulator	772	4	25.0	127	2	AS4670	RNA polymerase II
700	4	25.0	118	2	A87018	conserved hypotnet	773	4	25.0	127	2	A28084	DNA-binding protei
701	4	25.0	118	2	E70785	hypothetical prote	774	4	25.0	127	2	T37112	hypothetical protei
702	4	25.0	118	2	T35543	hypothetical prote	775	4	25.0	127	2	F82197	RatB2 protein VC14
703	4	25.0	118	2	T64028	hypothetical prote	776	4	25.0	128	2	C96509	protein F27F5.13 [
704	4	25.0	119	2	S75418	ribosomal protein	777	4	25.0	128	2	I51605	cholesterolkinin pr
705	4	25.0	119	2	S24294	chorion protein -	778	4	25.0	128	2	FN0647	epsN protein - Vib
706	4	25.0	119	2	S24292	chorion protein -	779	4	25.0	128	2	F71170	hypothetical prote
707	4	25.0	119	2	S24291	early chorion prot	780	4	25.0	128	2	B83204	lactoylgutathione
708	4	25.0	119	2	B45937	hypothetical prote	781	4	25.0	128	1	C71157	thioredoxin limpor
709	4	25.0	119	2	S76274	hypothetical prote	782	4	25.0	129	1	C71157	hypothetical prote
710	4	25.0	119	2	T48745	hypothetical prote	783	4	25.0	129	2	T06256	dormancy-associate
711	4	25.0	119	2	A23063	pancreatic B-cell	784	4	25.0	129	2	S28237	NADH2 dehydrogenas
712	4	25.0	119	2	E72637	hypothetical prote	785	4	25.0	129	2	B72287	conserved hypotnet
713	4	25.0	119	2	T18114	hypothetical prote	786	4	25.0	129	2	T21290	hypothetical prote
714	4	25.0	119	2	T39694	hypothetical prote	787	4	25.0	129	2	B69955	hypothetical prote
715	4	25.0	119	2	D87400	conserved hypotnet	788	4	25.0	129	2	F97776	hypothetical prote
716	4	25.0	119	2	CH4504	probable CCHC-type	789	4	25.0	129	2	C90314	hypothetical prote
717	4	25.0	120	1	MAWLB8	B4 protein - bovin	790	4	25.0	129	2	T05474	hypothetical prote
718	4	25.0	120	2	AH3515	sensory transducti	791	4	25.0	129	2	AD2700	conserved hypotnet
719	4	25.0	120	2	B89932	conserved hypotnet	792	4	25.0	130	2	Q7BP77	gene 7.7 protein -
720	4	25.0	120	2	I40821	regA protein - Clo	793	4	25.0	130	2	E56976	transfer complex p
721	4	25.0	120	2	C72234	hypothetical prote	794	4	25.0	130	2	D72780	hypothetical prote
722	4	25.0	120	2	T16442	hypothetical prote	795	4	25.0	130	2	T47291	hypothetical prote
723	4	25.0	120	2	F82153	hypothetical prote	796	4	25.0	130	2	E87024	probable membrane
724	4	25.0	120	2	H83602	hypothetical prote	797	4	25.0	130	2	PC4423	transcription regu
725	4	25.0	120	2	A99188	SSU ribosomal prot	798	4	25.0	130	2	H82909	DNA-binding protei
726	4	25.0	120	2	A87462	hypothetical prote	799	4	25.0	131	2	C75526	hypothetical prote
727	4	25.0	120	2	G97833	hypothetical prote	800	4	25.0	131	2	AG1371	hypothetical prote
728	4	25.0	121	2	S24293	chorion class CA p	801	4	25.0	131	2	S76123	lactoylgutathione
729	4	25.0	121	2	G98251	filx protein (AP27	802	4	25.0	131	2	UC7739	paralytic peptide
730	4	25.0	121	2	AE3034	chemotaxis respons	803	4	25.0	132	2	S73552	ribosomal protein
731	4	25.0	121	2	C64460	hypothetical prote	804	4	25.0	132	2	I46955	retinol-binding pr
732	4	25.0	121	2	I64187	opacity protein ho	805	4	25.0	132	2	S77014	transcription repr
733	4	25.0	121	2	CH7109	hypothetical prote	806	4	25.0	132	2	T25694	hypothetical prote
734	4	25.0	121	2	T04536	hypothetical prote	807	4	25.0	132	2	AC0215	hypothetical prote
735	4	25.0	121	2	F89970	conserved hypotnet	808	4	25.0	132	2	UC7173	5'-carboxymethyl-2-
736	4	25.0	121	2	AF1730	hypothetical prote	809	4	25.0	132	2	T32373	N,N-dimethylformam
737	4	25.0	122	2	T48951	hypothetical prote	810	4	25.0	132	2	D82954	hypothetical prote
738	4	25.0	122	2	C83866	hypothetical prote	811	4	25.0	132	2	AE2025	hypothetical prote
739	4	25.0	122	2	A83283	hypothetical prote	812	4	25.0	132	2	A87277	hypothetical prote
740	4	25.0	122	2	AC0423	conserved hypotnet	813	4	25.0	132	2	T12108	glycine-rich prote
741	4	25.0	123	2	AS8671	alpha-lactalbumin	814	4	25.0	132	2	E72491	hypothetical prote
742	4	25.0	123	2	I51604	cholecystokinin pr	815	4	25.0	132	2	AF2080	hypothetical prote
743	4	25.0	123	2	C30556	Ig heavy chain V r	816	4	25.0	133	1	PSOXG	phospholipase A2 (
744	4	25.0	123	2	AE0293	conserved hypotnet	817	4	25.0	133	2	T14810	ribosomal protein
745	4	25.0	123	2	F75293	conserved hypotnet	818	4	25.0	133	2	G75061	hypothetical prote
746	4	25.0	123	2	G90196	hypothetical prote	819	4	25.0	133	2	H90402	hypothetical prote
747	4	25.0	123	2	T30383	hypothetical prote	820	4	25.0	133	2	T04126	preproteine translo
748	4	25.0	123	2	H69972	hypothetical prote	821	4	25.0	133	2	AD1875	hypothetical prote
749	4	25.0	124	1	T1BHB	trypsin inhibitor	822	4	25.0	134	2	B35542	ribosomal protein
750	4	25.0	124	2	B55257	Ig gamma heavy cha	823	4	25.0	134	2	T28704	hypothetical prote
751	4	25.0	124	2	AE4391	hypothetical prote	824	4	25.0	134	2	A25748	hypothetical prote
752	4	25.0	124	2	T4876	hypothetical prote	825	4	25.0	134	2	B25748	lctB protein - Bac
753	4	25.0	124	2	B82196	hypothetical prote	826	4	25.0	134	2	S317736	hypothetical prote
754	4	25.0	124	2	G87326	RscB1 protein VC14	827	4	25.0	134	2	AE1741	hypothetical prote
755	4	25.0	124	2	A11067	probable membrane	828	4	25.0	134	2	D84402	hypothetical prote
756	4	25.0	124	2	B87345	hypothetical prote	829	4	25.0	134	2	G69947	phage-related prot
757	4	25.0	124	2	C82770	hypothetical prote	830	4	25.0	135	1	JU0139	interleukin-4 prec
758	4	25.0	125	2	B95116	hypothetical prote	831	4	25.0	135	2	H64481	hypothetical prote
759	4	25.0	125	2	C97986	large conductance	832	4	25.0	135	2	A75139	hypothetical prote

833	4	25.0	135	2	E83389	hypothetical prote	906	4	25.0	144	2	S01700	signal recognition
834	4	25.0	135	2	G90475	hypothetical prote	907	4	25.0	144	2	G70894	hypothetical prote
835	4	25.0	135	2	F86239	protein F20B24.4 l	908	4	25.0	144	2	D87208	conserved membrane
836	4	25.0	135	2	E84643	hypothetical prote	909	4	25.0	144	2	T18867	hypothetical prote
837	4	25.0	135	2	JC6313	lactoylgutathione	910	4	25.0	144	2	G71489	hypothetical prote
838	4	25.0	135	2	E70777	hypothetical prote	911	4	25.0	144	2	T33736	hypothetical prote
839	4	25.0	135	2	T21607	PMFamide-like pep	912	4	25.0	144	2	C95308	hypothetical prote
840	4	25.0	135	2	AC0695	lactoylgutathione	913	4	25.0	145	1	A35932	angiogenin precurs
841	4	25.0	135	2	D85772	lactoylgutathione	914	4	25.0	145	2	S60456	cysteine proteinas
842	4	25.0	135	2	H90923	lactoylgutathione	915	4	25.0	145	2	B69158	sensory transducti
843	4	25.0	135	2	A82607	conserved hypotnet	916	4	25.0	145	2	F75337	transposase - Dein
844	4	25.0	135	2	E64922	lactoylgutathione	917	4	25.0	145	2	G71024	hypothetical prote
845	4	25.0	136	2	F87461	hypothetical prote	918	4	25.0	145	2	S23005	tram protein - Bsc
846	4	25.0	136	2	F81970	hypothetical prote	919	4	25.0	145	2	H84174	hypothetical prote
847	4	25.0	136	2	T37074	hypothetical prote	920	4	25.0	145	2	F82189	hypothetical prote
848	4	25.0	137	2	T14828	H+-transporting tw	921	4	25.0	145	2	G90816	hypothetical prote
849	4	25.0	137	2	D81914	probable membrane	922	4	25.0	145	2	B96705	unknown protein, 8
850	4	25.0	137	2	E81187	conserved hypotnet	923	4	25.0	145	2	E84699	hypothetical prote
851	4	25.0	137	2	T21378	hypothetical prote	924	4	25.0	145	2	T52471	hypothetical prote
852	4	25.0	137	2	C84689	hypothetical prote	925	4	25.0	146	1	G71158	hypothetical prote
853	4	25.0	137	2	H84345	inosine-5'-monopho	926	4	25.0	146	2	E97249	probable acetyltra
854	4	25.0	137	2	A83202	hypothetical prote	927	4	25.0	146	2	S54169	glycine rich prote
855	4	25.0	137	2	T15084	hypothetical prote	928	4	25.0	146	2	T08537	tram protein - Ent
856	4	25.0	137	2	T21608	hypothetical prote	929	4	25.0	146	2	S73456	DNA-directed RNA p
857	4	25.0	138	2	T07385	hypothetical prote	930	4	25.0	146	2	AC2423	hypothetical prote
858	4	25.0	138	2	C90218	14-3-3 protein tft	931	4	25.0	146	2	B69179	hypothetical prote
859	4	25.0	138	2	AB0538	15U ribosomal prot	932	4	25.0	147	1	NRHUNG	angiogenin precurs
860	4	25.0	138	2	G81211	probable secreted	933	4	25.0	147	1	FZSVN2	nodulin-24 - soybe
861	4	25.0	138	2	B64403	lactoylgutathione	934	4	25.0	147	2	F81286	30S ribosomal prot
862	4	25.0	138	2	AH2106	hypothetical prote	935	4	25.0	147	2	AD0053	ribosomal-protein-
863	4	25.0	139	2	A41404	hypothetical prote	936	4	25.0	147	2	S65579	hypothetical prote
864	4	25.0	139	2	I46630	cathepsin L (EC 3.	937	4	25.0	147	2	G71282	hypothetical prote
865	4	25.0	139	2	S54233	rearranged T-cell	938	4	25.0	147	2	B96634	unknown protein, 4
866	4	25.0	139	2	G84000	Ig mu heavy chain	939	4	25.0	147	2	H87574	hbaa protein [limp
867	4	25.0	139	2	F72672	transcription regu	940	4	25.0	147	2	A42610	holdfast protein H
868	4	25.0	139	2	T03184	hypothetical prote	941	4	25.0	147	2	S28698	hypothetical prote
869	4	25.0	139	2	T50947	probable high-mobi	942	4	25.0	147	2	JQ1320	high mobility grou
870	4	25.0	139	2	C87459	HSP10 related prot	943	4	25.0	148	1	HSUR8P	histone H2B.3, spe
871	4	25.0	139	2	B75522	hypothetical prote	944	4	25.0	148	2	TE0320	ribosomal protein
872	4	25.0	139	2	T73968	hypothetical prote	945	4	25.0	148	2	G87362	transcription regu
873	4	25.0	140	1	S14320	hypothetical prote	946	4	25.0	148	2	S77438	hypothetical prote
874	4	25.0	140	1	C95999	alpha-atrial natrri	947	4	25.0	148	2	AD2293	mutator protein [i
875	4	25.0	140	2	B86739	hypothetical prote	948	4	25.0	148	2	E90359	partial transposas
876	4	25.0	141	2	S35610	potassium channel	949	4	25.0	148	2	C88922	protein W03P9.5 [i
877	4	25.0	141	2	S35611	cytochrome P450 4A	950	4	25.0	149	2	B82543	50S ribosomal prot
878	4	25.0	141	2	A48583	cytochrome P450 4A	951	4	25.0	149	2	T37099	hypothetical prote
879	4	25.0	141	2	S15145	ribosomal protein	952	4	25.0	149	2	AC0375	hypothetical prote
880	4	25.0	141	2	G95996	hypothetical prote	953	4	25.0	149	2	D97333	hypothetical prote
881	4	25.0	141	2	B82461	conserved hypotnet	954	4	25.0	149	2	E90230	partial transposas
882	4	25.0	141	2	H82462	hypothetical prote	955	4	25.0	150	2	T25581	hypothetical prote
883	4	25.0	141	2	E90519	hypothetical prote	956	4	25.0	150	2	S62376	aspartic proteinas
884	4	25.0	141	2	B1820	hypothetical prote	957	4	25.0	150	2	F84976	3-dehydroquinatase
885	4	25.0	141	2	C69339	potassium channel	958	4	25.0	150	2	S39980	hemoglobin II alph
886	4	25.0	141	2	C83768	hypothetical prote	959	4	25.0	150	2	T14457	MADS box protein h
887	4	25.0	142	1	E64794	hypotnetical prote	960	4	25.0	150	2	T37513	N-acetyltransferas
888	4	25.0	142	1	S06630	hyperglycemic horm	961	4	25.0	150	2	T39482	T-cadherin
889	4	25.0	142	2	S12311	glycine-rich RNA-b	962	4	25.0	150	2	S55903	phosphotransferase
890	4	25.0	142	2	A10577	conserved hypotnet	963	4	25.0	150	2	E64976	pts system, galact
891	4	25.0	142	2	B85560	hypothetical prote	964	4	25.0	150	2	A90991	pts system, galact
892	4	25.0	142	2	F90709	hypothetical prote	965	4	25.0	150	2	C85836	pts system, galact
893	4	25.0	143	1	HSUR6P	histone H2B.2, spe	966	4	25.0	150	2	S66926	hypothetical prote
894	4	25.0	143	2	A82216	cytosine deaminase	967	4	25.0	150	2	H97466	hypothetical prote
895	4	25.0	143	2	E72504	hypotnetical prote	968	4	25.0	150	2	AC2685	conserved hypotnet
896	4	25.0	143	2	C32106	glucan endo-1,3-de	969	4	25.0	150	2	D97627	probable acetyltra
897	4	25.0	143	2	AC2289	mutator protein [i	970	4	25.0	150	2	AF2850	acetyltransferase
898	4	25.0	143	2	DB1751	conserved hypotnet	971	4	25.0	150	2	S48926	hypothetical prote
899	4	25.0	143	2	H70601	hypothetical prote	972	4	25.0	151	2	PS0168	nucleocapsid prote
900	4	25.0	143	2	T14664	hypothetical prote	973	4	25.0	151	2	PS0173	nucleocapsid prote
901	4	25.0	143	2	S05717	hypothetical prote	974	4	25.0	151	2	PS0170	nucleocapsid prote
902	4	25.0	144	1	HSURB1	histone H2B.2, spe	975	4	25.0	151	2	PS0175	nucleocapsid prote
903	4	25.0	144	1	F87538	ExoD/Toir family p	976	4	25.0	151	2	PS0176	nucleocapsid prote
904	4	25.0	144	2	E97511	hypothetical prote	977	4	25.0	151	2	PS0169	nucleocapsid prote
905	4	25.0	144	2	A82730	conserved hypotnet	978	4	25.0	151	2	PS0171	nucleocapsid prote

579 4 25.0 151 2 E60174 nucleocapsid prote  
580 4 25.0 151 2 S52879 nucleocapsid prote  
581 4 25.0 151 2 S52884 nucleocapsid prote  
582 4 25.0 151 2 S52881 nucleocapsid prote  
583 4 25.0 151 2 S52880 nucleocapsid prote  
584 4 25.0 151 2 S52882 nucleocapsid prote  
585 4 25.0 151 2 S52873 nucleocapsid prote  
586 4 25.0 151 2 S52883 nucleocapsid prote  
587 4 25.0 151 2 S52875 nucleocapsid prote  
588 4 25.0 151 2 S52869 nucleocapsid prote  
589 4 25.0 151 2 S52870 nucleocapsid prote  
590 4 25.0 151 2 S75406 hypothetical prote  
591 4 25.0 151 2 A60931 probable exported  
592 4 25.0 151 2 PC4223 rho kinase - pig (  
593 4 25.0 151 2 T20071 hypothetical prote  
594 4 25.0 152 2 S57425 cysteine proteinas  
595 4 25.0 152 2 S57427 cysteine proteinas  
596 4 25.0 152 2 A32669 vasotocin 1 / neur  
597 4 25.0 152 2 H87244 50S ribosomal prot  
598 4 25.0 152 2 H70913 probable ribosomal  
599 4 25.0 152 2 T41567 hypothetical prote  
1000 4 25.0 152 2 T19905 hypothetical prote

## ALIGNMENTS

RESULT 1  
A28942  
pancreatic elastase (EC 3.4.21.36) precursor - fluke (Schistosoma mansoni)  
C:Species: Schistosoma mansoni  
C:Date: 30-Jun-1989 #sequence\_revision 30-Jun-1989 #text\_change 26-Aug-1999  
R:Newport, G.R.; McKerron, J.H.; Hedstrom, R.; Pettit, M.; McGarrigle, L.; Barr, P.J.; A  
J. Biol. Chem. 263, 13179-13184, 1988  
A:Title: Cloning of the proteinase that facilitates infection by schistosome parasites.  
A:Reference number: A28942; MUID:88330818; PMID:3166457  
A:Accession: A28942  
A:Molecule type: mRNA  
A:Residues: 1-264 <NEW>  
A:Cross-references: GB:J03946; NID:G160947; PID:G160948  
C:Superfamily: trypsin; trypsin homology  
C:Keywords: hydrolase; serine proteinase  
F:27-256/Domain: trypsin homology <TRY>  
Query Match 100.0%; Score 16; DB 2; Length 264;  
Best Local Similarity 100.0%; Pred. No. 9.4e-11;  
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 VGYGRDDNDSDSRKN 16  
Db 163 VGYGRDDNDSDSRKN 178  
RESULT 2  
E87490  
NADH dehydrogenase I, G subunit CC1946 [imported] - Caulobacter crescentus  
C:Species: Caulobacter crescentus  
C:Date: 20-Apr-2001 #sequence\_revision 20-Apr-2001 #text\_change 10-May-2001  
R:Nierman, W.C.; Feldblyum, T.V.; Paulsen, I.T.; Nelson, K.E.; Eisen, J.; Heidelberg, J.  
B.; Lab, M.T.; DeBoy, R.T.; Dodson, R.J.; Durkin, A.S.; Gwinn, M.L.; Haft, D.H.; Kjol  
n, J.; Ermolaeva, M.; White, O.; Salzberg, S.L.; Shapiro, L.; Venter, J.C.; Fraser, C.M.  
Proc. Natl. Acad. Sci. U.S.A. 98, 4136-4141, 2001  
A:Title: Complete Genome Sequence of Caulobacter crescentus.  
A:Reference number: A87249; MUID:21173698; PMID:11259647  
A:Accession: E87490  
A:Status: Preliminary  
A:Molecule type: DNA  
A:Residues: 1-686 <STO>  
A:Cross-references: GB:AE005673; NID:G13423405; PIDN:AAK23921.1; GSPDB:GN00148  
C:Genetics:  
A:Gene: CC1946

C:Superfamily: NADH dehydrogenase (ubiquinone) chain 11

Query Match 43.8%; Score 7; DB 2; Length 686;  
Best Local Similarity 100.0%; Pred. No. 2.8;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VGYGRDD 7  
Db 115 VGYGRDD 121

## RESULT 3

D69543  
hypothetical protein AF2348 - Archaeoglobus fulgidus

C:Species: Archaeoglobus fulgidus

C:Date: 05-Dec-1997 #sequence\_revision 05-Dec-1997 #text\_change 22-Oct-1999

A:Accession: D69543

R:Klenk, H.P.; Clayton, R.A.; Tomb, J.F.; White, O.; Nelson, K.E.; Ketchum, K.A.; Dodso  
n, J.; Fleischmann, R.D.; Quackenbush, J.; Lee, N.H.; Sutton, G.G.; Gill, S.; Kiriess, E.F

Glodek, A.; Zhou, L.; Overbeek, R.; Gocayne, J.D.; Weidman, J.F.; McDonald, L.

Nature 390, 364-370, 1997

A:Authors: Uterback, T.; Cotton, M.D.; Spriggs, T.; Artlich, P.; Kaine, B.P.; Sykes, S

Smith, H.O.; Woese, C.R.; Venter, J.C.

A:Title: The complete genome sequence of the hyperthermophilic, sulfate-reducing archae

A:Reference number: A69250; MUID:98049343; PMID:9389475

A:Accession: D69543

A:Status: preliminary; nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-112 <KLE>

A:Cross-references: GB:AE001114; GB:AE000782; NID:G2669437; PIDN:AA91315.1; PID:G26507

C:Superfamily: Archaeoglobus fulgidus hypothetical protein AF2348

Query Match 37.5%; Score 6; DB 2; Length 112;  
Best Local Similarity 100.0%; Pred. No. 6.9;

Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 11 DPSRKN 16  
Db 80 DPSRKN 85

## RESULT 4

G69532  
conserved hypothetical protein AF2263 - Archaeoglobus fulgidus

C:Species: Archaeoglobus fulgidus

C:Date: 05-Dec-1997 #sequence\_revision 05-Dec-1997 #text\_change 24-Nov-1999

A:Accession: G69532

R:Klenk, H.P.; Clayton, R.A.; Tomb, J.F.; White, O.; Nelson, K.E.; Ketchum, K.A.; Dodso

n, J.; Fleischmann, R.D.; Quackenbush, J.; Lee, N.H.; Sutton, G.G.; Gill, S.; Kiriess, E.F

Glodek, A.; Zhou, L.; Overbeek, R.; Gocayne, J.D.; Weidman, J.F.; McDonald, L.

Nature 390, 364-370, 1997

A:Authors: Uterback, T.; Cotton, M.D.; Spriggs, T.; Artlich, P.; Kaine, B.P.; Sykes, S

Smith, H.O.; Woese, C.R.; Venter, J.C.

A:Title: The complete genome sequence of the hyperthermophilic, sulfate-reducing archae

A:Reference number: A69250; MUID:98049343; PMID:9389475

A:Accession: G69532

A:Status: preliminary; nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-122 <KLE>

A:Cross-references: GB:AE000948; GB:AE000782; NID:G2669271; PIDN:AA88987.1; PID:G26482

C:Superfamily: Archaeoglobus fulgidus conserved hypothetical protein AF2263

Query Match 37.5%; Score 6; DB 2; Length 122;  
Best Local Similarity 100.0%; Pred. No. 7.5;

Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 GYGRDD 7  
Db 28 GYGRDD 33

## RESULT 5

AG0689

alternative bacteriophage tail fiber C-terminus [imported] - *Salmonella enterica* subsp.  
C/Species: *Salmonella enterica* subsp. *enterica* serovar Typh  
A/Note: this species has also been called *Salmonella typhi*  
C/Date: 09-Nov-2001 #sequence\_revision 09-Nov-2001 #text\_change 17-May-2002  
C/Accession: AG0689  
R/Parikh, U.; Dugan, G.; James, K.D.; Thomson, N.R.; Pickard, D.; Main, J.; Churcher,  
th, T.; Conerton, P.; Cronin, A.; Davis, P.; Davies, R.M.; Dowd, L.; White, N.; Farrar,  
S.; Moul, S.; O'Gaora, P.  
Nature 413, 848-852, 2001  
A/Authors: Parry, C.; Quail, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.;  
A/Title: Complete genome sequence of a multiple drug resistant *Salmonella enterica* serov  
A/Reference number: AB0502; PMID:11677608  
A/Accession: AG0689  
A/Status: preliminary  
A/Molecule type: DNA  
A/Residues: 1-160 <PAR>  
A/Cross-references: GB:AL513382; PIDN:CAD01888.1; PID:g16502731; GSPDB:GN00176  
C/Genetics:  
A/Gene: STY1642  
C/Superfamily: phage P1 major tail fiber protein S

Query Match 37.5%; Score 6; DB 2; Length 160;  
Best Local Similarity 100.0%; Pred. No. 9.6;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 4 GRDND 9  
Db 86 GRDND 91

RESULT 6  
S74031  
hypothetical protein c0213 - *Sulfolobus solfataricus*  
C/Species: *Sulfolobus solfataricus*  
C/Date: 09-Oct-1997 #sequence\_revision 24-Oct-1997 #text\_change 17-Mar-1999  
C/Accession: S74031  
R/Sensen, C.W.; Klenk, H.P.; Singh, R.K.; Allard, G.; Chan, C.C.Y.; Liu, Q.Y.; Penny, S.  
Mol. Microbiol. 22, 175-191, 1996  
A/Title: Organizational characteristics and information content of an archaeal genome: 1  
A/Reference number: S73076; MUID:97055432; PMID:8899719  
A/Accession: S74031  
A/Status: nucleic acid sequence not shown; translation not shown  
A/Molecule type: DNA  
A/Residues: 1-243 <SEN>  
A/Cross-references: EMBL:X08256; NID:g1707679; PID:e283907; PID:g1707725  
A/Experimental source: strain P2  
A/Note: the nucleotide sequence was submitted to the EMBL Data Library, September 1996

Query Match 37.5%; Score 6; DB 2; Length 243;  
Best Local Similarity 100.0%; Pred. No. 14;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 4 GRDND 9  
Db 216 GRDND 221

RESULT 7  
A98263  
transcription regulatory protein tccD [imported] - *Agrobacterium tumefaciens* (strain C58  
C/Species: *Agrobacterium tumefaciens*  
C/Date: 22-Oct-2001 #sequence\_revision 22-Oct-2001 #text\_change 11-Jan-2002  
C/Accession: A98263  
R/Goodner, B.; Hinkle, G.; Gattung, S.; Miller, N.; Blanchard, M.; Qurollo, B.; Goldman,  
A.; Liu, F.; Wollam, C.; Allinger, M.; Dougherty, D.; Scott, C.; Lappas, C.; Markelz, B.;  
Science 294, 2323-2328, 2001  
A/Title: Genome Sequence of the Plant Pathogen and Biotechnology Agent *Agrobacterium tum*  
A/Reference number: A97359; PMID:11743194  
A/Accession: A98263  
A/Status: preliminary  
A/Molecule type: DNA  
A/Residues: 1-244 <KUR>  
A/Cross-references: GB:AE007870; PIDN:AAK89627.1; PID:g15159523; GSPDB:GN00170

C/Genetics:  
A/Gene: AGR L 2108  
A/Map position: linear chromosome  
C/Superfamily: ompR protein, response regulator homology

Query Match 37.5%; Score 6; DB 2; Length 244;  
Best Local Similarity 100.0%; Pred. No. 14;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2 GYGRD 7  
Db 14 GYGRD 19

RESULT 8  
AF3032  
conserved hypothetical protein Atu3867 [imported] - *Agrobacterium tumefaciens* (strain C  
C/Species: *Agrobacterium tumefaciens*  
C/Date: 11-Jan-2002 #sequence\_revision 11-Jan-2002 #text\_change 11-Jan-2002  
C/Accession: AF3032  
R/Mood, D.W.; Sebhal, J.C.; Kaul, R.; Monke, D.; Chen, L.; Wood, G.E.; Chen, Y.; Woo,  
erage, G.; Gillet, W.; Grant, C.; Guenther, D.; Kutayvin, T.; Levy, R.; Li, M.; McClel  
e, Karp, P.; Romero, P.; Zhang, S.  
Science 294, 2317-2323, 2001  
A/Authors: Yoo, H.; Tao, Y.; Biddle, P.; Jung, M.; Krespan, W.; Perry, M.; Gordon-Kamm,  
ster, E.W.  
A/Title: The Genome of the Natural Genetic Engineer *Agrobacterium tumefaciens* C58.  
A/Reference number: AB2577; PMID:11743193  
A/Accession: AF3032  
A/Status: preliminary  
A/Molecule type: DNA  
A/Residues: 1-292 <KUR>  
A/Cross-references: GB:AE008689; PIDN:AL44676.1; PID:g17742303; GSPDB:GN00187  
C/Genetics:  
A/Experimental source: strain C58 (Dupont)  
A/Gene: Atu3867  
A/Map position: linear chromosome

Query Match 37.5%; Score 6; DB 2; Length 292;  
Best Local Similarity 100.0%; Pred. No. 17;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 VGYGRD 6  
Db 258 VGYGRD 263

RESULT 9  
F98253  
hypothetical protein AGR L 1957 [imported] - *Agrobacterium tumefaciens* (strain C58, Cer  
C/Species: *Agrobacterium tumefaciens*  
C/Date: 22-Oct-2001 #sequence\_revision 22-Oct-2001 #text\_change 11-Jan-2002  
C/Accession: F98253  
R/Goodner, B.; Hinkle, G.; Gattung, S.; Miller, N.; Blanchard, M.; Qurollo, B.; Goldman,  
A.; Liu, F.; Wollam, C.; Allinger, M.; Dougherty, D.; Scott, C.; Lappas, C.; Markelz, B.  
Science 294, 2323-2328, 2001  
A/Title: Genome Sequence of the Plant Pathogen and Biotechnology Agent *Agrobacterium tum*  
A/Reference number: A97359; PMID:11743194  
A/Accession: F98253  
A/Status: preliminary  
A/Molecule type: DNA  
A/Residues: 1-303 <KUR>  
A/Cross-references: GB:AE007870; PIDN:AAK89552.1; PID:g15159434; GSPDB:GN00170  
C/Genetics:  
A/Gene: AGR L 1957  
A/Map position: linear chromosome

Query Match 37.5%; Score 6; DB 2; Length 303;  
Best Local Similarity 100.0%; Pred. No. 17;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 VGYGRD 6  
Db 258 VGYGRD 263

Db 269 VGYGRD 274

RESULT 10

AB2193 hypothetical protein alr1097 [imported] - Nostoc sp. (strain PCC 7120)

C/Species: Nostoc sp.

A/Note: Nostoc sp. strain PCC 7120 is a synonym of Anabaena sp. strain PCC 7120

C/Date: 14-Dec-2001 #sequence\_revision 14-Dec-2001 #text\_change 30-Jun-2002

C/Accession: AB2193

R/Kaneko, T.; Nakamura, Y.; Molk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.; Iriyuchi, N.; Shimpo, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Yasuda, M.; Tabata, S.

DNA Res. 8, 205-213, 2001

A/Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium Anabaena

A/Reference number: AB1807; MUID:21595285; PMID:11759840

A/Accession: AB2193

A/Status: preliminary

A/Molecule type: DNA

A/Residues: 1-935 <KUR>

A/Cross-references: GB:BA000019; PIDN:BA874796.1; PID:G17132191; GSPDB:GN00179

A/Experimental source: strain PCC 7120

C/Genetics:

A/Gene: alr3097

Query Match 37.5%; Score 6; DB 2; Length 335;  
Best Local Similarity 100.0%; Pred. No. 19;

Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 VGYGRD 6  
|||||

Db 169 VGYGRD 174

RESULT 11

A70839 Probable membrane protein - Mycobacterium tuberculosis (strain H37RV)

N/Alternate names: membrane protein MTCu622 homolog

C/Species: Mycobacterium tuberculosis

C/Date: 17-Jul-1998 #sequence\_revision 17-Jul-1998 #text\_change 22-Oct-1999

C/Accession: A70839

R/Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.; Connor, R.; Davies, R.; Devlin, K.; Felwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.

Rajadurai, M.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.

Nature 393, 537-544, 1998

A/Autours: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.

A/Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome

A/Reference number: A70500; MUID:98295987; PMID:9634230

A/Accession: A70839

A/Status: preliminary; nucleic acid sequence not shown; translation not shown

A/Molecule type: DNA

A/Residues: 1-412 <COB>

A/Cross-references: GB:AL021928; GB:AL123456; NID:G3261522; PIDN:CAA17326.1; PID:el25242

A/Experimental source: strain H37RV

C/Genetics:

A/Gene: RV0204c

Query Match 37.5%; Score 6; DB 2; Length 412;  
Best Local Similarity 100.0%; Pred. No. 23;

Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 7 DNDRDP 12  
|||||

Db 363 DNDRDP 368

RESULT 12

S73419

signal recognition particle protein f5h - Mycoplasma pneumoniae (strain ATCC 29342)

N/Alternate names: hypothetical protein D09\_orf450

C/Species: Mycoplasma pneumoniae

A/Variety: ATCC 29342

C/Date: 26-Feb-1997 #sequence\_revision 25-Apr-1997 #text\_change 07-Dec-1999

C/Accession: S73419

R/Himmelreich, R.; Hilbert, H.; Plagens, H.; Pirk, E.; Li, B.C.; Herrmann, R.

Nucleic Acids Res. 24, 4420-4449, 1996

A/Title: Complete sequence analysis of the genome of the bacterium Mycoplasma pneumonia

A/Reference number: S73327; MUID:97105885; PMID:8948633

A/Accession: S73419

A/Status: preliminary; nucleic acid sequence not shown; translation not shown

A/Molecule type: DNA

A/Residues: 1-450 <HIM>

A/Cross-references: EMBL:AE000011; GB:U00089; NID:G1673740; PIDN:AAB95741.1; PID:G16737

A/Note: the nucleotide sequence was submitted to the EMBL Data Library, November 1996

C/Genetics:

A/Gene: f5h

A/Gene: f5h

C/Superfamily: signal recognition particle 54k protein

Query Match 37.5%; Score 6; DB 2; Length 450;  
Best Local Similarity 100.0%; Pred. No. 25;

Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 10 RDPNRK 15  
|||||

Db 392 RDPNRK 397

RESULT 13

S74935 hypothetical protein sl10641 - Synechocystis sp. (strain PCC 6803)

C/Species: Synechocystis sp.

A/Variety: PCC 6803

C/Date: 25-Apr-1997 #sequence\_revision 25-Apr-1997 #text\_change 08-Oct-1999

C/Accession: S74935

R/Kaneko, T.; Sato, S.; Kotani, H.; Tanaka, A.; Asamizu, E.; Nakamura, Y.; Miyajima, N.

O, K.; Okumura, S.; Shimpo, S.; Takeuchi, C.; Wada, T.; Watanabe, A.; Yamada, M.; Yasud

DNA Res. 3, 109-136, 1996

A/Title: Sequence analysis of the genome of the unicellular cyanobacterium Synechocystis

A/Reference number: S74322; MUID:97061201; PMID:8905231

A/Accession: S74935

A/Status: preliminary

A/Molecule type: DNA

A/Residues: 1-509 <KAN>

A/Cross-references: EMBL:D90902; GB:AB001339; NID:G1652027; PIDN:BA16975.1; PID:dl0177

A/Note: the nucleotide sequence was submitted to the EMBL Data Library, June 1996

Query Match 37.5%; Score 6; DB 2; Length 509;  
Best Local Similarity 100.0%; Pred. No. 28;

Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 VGYGRD 6  
|||||

Db 376 VGYGRD 381

RESULT 14

S63208

hypothetical protein YNL242w - yeast (Saccharomyces cerevisiae)

N/Alternate names: hypothetical protein N1106

C/Species: Saccharomyces cerevisiae

C/Date: 27-Apr-1996 #sequence\_revision 03-May-1996 #text\_change 19-Apr-2002

C/Accession: S63208; S63210; S63187; S67356; S72074

R/Pandolfo, D.; De Antoni, A.; Lantranchi, G.; Valle, G.

submitted to the Protein Sequence Database, April 1996

A/Reference number: S63188

A/Accession: S63208

A/Molecule type: DNA

A/Residues: 1-1592 <PAN>

A/Cross-references: EMBL:Z71518; NID:G1302277; PID:e239636; PID:G1302278; MIPS:YNL242w

A/Experimental source: strain S288C

R/Satz, J.E.; Baladron, V.; Del Rey, F.

submitted to the Protein Sequence Database, April 1996

A/Reference number: S63210

A/Accession: S63210

A/Molecule type: DNA



A;Residues: 1-277 <SAI>  
A;Cross-references: EMBL:Z71518; MIPS:YNL242w  
A;Experimental source: strain S288C  
R;Garcia-Cantalejo, J.M.; Boskovic, J.; Jimenez, A.  
Submitted to the Protein Sequence Database, April 1996  
A;Reference number: S63030  
A;Accession: S63187  
A;Molecule type: DNA  
A;Residues: 278-1304 <GAR>  
A;Cross-references: EMBL:Z71518; MIPS:YNL242w  
A;Experimental source: strain S288C  
R;Pandofo, D.; de Antoni, A.; Lanfranchi, G.; Valle, G.  
Submitted to the EMBL Data Library, February 1996  
A;Description: DNA sequence of cosmid 14-5 from chromosome XIV.  
A;Reference number: S67355  
A;Accession: S67356  
A;Molecule type: DNA  
A;Residues: 1-1592 <PAW>  
A;Cross-references: EMBL:Z69381; NID:G1183970; PID:G1183972  
R;Pandofo, D.; de Antoni, A.; Lanfranchi, G.; Valle, G.  
Yeast 12, 1071-1076, 1996  
A;Title: The DNA sequence of cosmid 14-5 from chromosome XIV reveals 21 open reading frames  
A;Reference number: S72073; MUID:97051596; PMID:8896273  
A;Accession: S72074  
A;Status: nucleic acid sequence not shown; translation not shown  
A;Molecule type: DNA  
A;Residues: 1-1592 <PAF>  
A;Cross-references: EMBL:Z69381; NID:G1183970; PID:G1183972  
A;Note: the nucleotide sequence was submitted to the EMBL Data Library, February 1996  
C;Genetics:  
A;Gene: SGD:APG2  
A;Cross-references: SGD:S0005186  
A;Map position: 14L  
A;Note: YNL242w

Query Match 37.5%; Score 6; DB 2; Length 1592;  
Best Local Similarity 100.0%; Pred. No. 80;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 8 NDRDPS 13  
Db 190 NDRDPS 195

RESULT 15  
S24713  
Ig alpha chain, truncated (MEC) - human  
C;Species: Homo sapiens (man)  
C;Date: 20-Feb-1995 #sequence\_revision 20-Feb-1995 #text\_change 16-Aug-1996  
C;Accession: S24713  
R;Teapla, A.  
Submitted to the EMBL Data Library, August 1992  
A;Reference number: S24708  
A;Accession: S24713  
A;Status: preliminary  
A;Molecule type: mRNA  
A;Residues: 1-45 <TSA>  
A;Cross-references: EMBL:Z14964  
C;Keywords: Immunoglobulin

Query Match 31.2%; Score 5; DB 2; Length 45;  
Best Local Similarity 100.0%; Pred. No. 39;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 9 DRDPS 13  
Db 24 DRDPS 28

RESULT 16  
G90800  
hypoetical protein Ec61375 [imported] - Escherichia coli (strain O157:H7, substrain R1  
C;Species: Escherichia coli

C;Date: 18-Jul-2001 #sequence\_revision 18-Jul-2001 #text\_change 18-Jul-2001  
C;Accession: G90800  
R;Hayashi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.C.  
gasawara, N.; Yasunaga, T.; Kuhnara, S.; Shiba, T.; Hattori, M.; Shinagawa, H.  
DNA Res. 8, 11-22, 2001  
A;Title: Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and gen  
A;Reference number: A99629; MUID:21156231; PMID:11258796  
A;Accession: G90800  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-63 <HAY>  
A;Cross-references: GB:BA000007; PIDN:BA934798.1; PID:G13360835; GSPDB:GN00154  
A;Experimental source: strain O157:H7, substrain R1MD 0509952  
C;Genetics:  
A;Gene: Ec61375

Query Match 31.2%; Score 5; DB 2; Length 63;  
Best Local Similarity 100.0%; Pred. No. 54;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 7 DNRD 11  
Db 12 DNRD 16

RESULT 17  
D81239  
hypoetical protein NMB0092 [imported] - Neisseria meningitidis (strain MCS8 serogroup  
C;Species: Neisseria meningitidis  
C;Date: 31-Mar-2000 #sequence\_revision 31-Mar-2000 #text\_change 19-Jan-2001  
C;Accession: D81239  
R;Tettelin, H.; Saunders, N.J.; Heidelberg, J.; Jeffries, A.C.; Nelson, K.E.; Eisen, J.  
Hickey, E.K.; Haft, D.H.; Salzberg, S.L.; White, O.; Fleischmann, R.D.; Dougherty, B.A.  
ri, H.; Qin, H.; Vamathevan, J.; Gill, J.; Scarlato, V.; Massignani, V.; Pizze, M.  
Science 287, 1809-1815, 2000  
A;Authors: Grandi, G.; Sun, L.; Smith, H.O.; Fraser, C.M.; Moxon, E.R.; Rappuoli, R.; V  
A;Title: Complete genome sequence of Neisseria meningitidis serogroup B strain MCS8.  
A;Reference number: A81000; MUID:20175755; PMID:10710307  
A;Accession: D81239  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-75 <TET>  
A;Cross-references: GB:AE002368; GB:AE002098; NID:G7225303; PIDN:AAF40554.1; PID:G72253  
A;Experimental source: serogroup B, strain MCS8  
C;Genetics:  
A;Gene: NMB0092

Query Match 31.2%; Score 5; DB 2; Length 75;  
Best Local Similarity 100.0%; Pred. No. 63;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 12 PSRXN 16  
Db 54 PSRXN 58

RESULT 18  
H72412  
hypoetical protein - Thermotoga maritima (strain MSB8)  
C;Species: Thermotoga maritima  
C;Date: 11-Jun-1999 #sequence\_revision 11-Jun-1999 #text\_change 21-Jul-2000  
C;Accession: H72412  
R;Nelson, K.E.; Clayton, R.A.; Gill, S.R.; Gwinn, M.L.; Dodson, R.J.; Haft, D.H.; Hicke  
Garrett, M.W.; Stewart, A.M.; Cotton, M.D.; Pratta, M.S.; Phillips, C.A.; Richardson, D.  
C.M.  
Nature 399, 323-329, 1999  
A;Title: Evidence for lateral gene transfer between Archaea and Bacteria from genome se  
A;Reference number: A72200; MUID:99287316; PMID:10360571  
A;Accession: H72412  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-76 <ARN>  
A;Cross-references: GB:AE001700; GB:AE000512; NID:G4980636; PIDN:AMD35247.1; PID:G49806

A:Experimental source: strain MSB8  
C:Genetics:  
A:Gene: TM0154

Query Match 31.2%: Score 5; DB 2; Length 76;  
Best Local Similarity 100.0%; Pred. No. 64;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 12 PSRKN 16  
|||||  
Db 27 PSRKN 31

RESULT 19  
A91036  
hypothetical protein ECa3257 [imported] - Escherichia coli (strain O157:H7, substrain R1  
C:Species: Escherichia coli  
C:Date: 18-Jul-2001 #sequence\_revision 18-Jul-2001 #text\_change 18-Jul-2001  
C:Accession: A91036  
R:Hayashi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.G.  
gasawara, N.; Yasunaga, T.; Kuhara, S.; Shiba, T.; Hattori, M.; Shinagawa, H.  
DNA Res. 8, 11-22, 2001  
A:Title: Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and gene  
A:Reference number: A99629; MUID:21156231; PMID:11258796  
A:Accession: A91036  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-80 <HAY>  
A:Cross-references: GB:BA000007; PIDN:BA936680.1; PID:813362727; GSPDB:GN00154  
A:Experimental source: strain O157:H7, substrain R1MD 050952  
C:Genetics:  
A:Gene: ECa3257

Query Match 31.2%: Score 5; DB 2; Length 80;  
Best Local Similarity 100.0%; Pred. No. 67;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VGYGR 5  
|||||  
Db 75 VGYGR 79

RESULT 20  
C85880  
hypothetical protein Z3642 [imported] - Escherichia coli (strain O157:H7, substrain EDLg  
C:Species: Escherichia coli  
C:Date: 16-Feb-2001 #sequence\_revision 16-Feb-2001 #text\_change 14-Sep-2001  
C:Accession: C85880  
R:Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glaesner, J.D.; Rose, D.J.; Mayhew  
Miller, L.; Grobeck, E.J.; Davis, N.W.; Lim, A.; Dimalanta, E.; Potamouls, K.; Apodaca,  
Nature 409, 529-533, 2001  
A:Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.  
A:Reference number: A85480; MUID:21074935; PMID:11206551  
A:Accession: C85880  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-80 <STO>  
A:Cross-references: GB:AE005174; NID:g12516751; PIDN:AAG57503.1; GSPDB:GN00145; UWGP:Z36  
C:Genetics:  
A:Experimental source: strain O157:H7, substrain EDL933  
A:Gene: Z3642

Query Match 31.2%: Score 5; DB 2; Length 80;  
Best Local Similarity 100.0%; Pred. No. 67;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VGYGR 5  
|||||  
Db 75 VGYGR 79

RESULT 21  
F65011

hypothetical protein b2377 - Escherichia coli (strain K-12)  
C:Species: Escherichia coli  
C:Date: 12-Sep-1997 #sequence\_revision 17-Sep-1997 #text\_change 01-Mar-2002  
C:Accession: F65011  
R:Battner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.; C  
A.; Rose, D.J.; Mau, B.; Shao, Y.  
Science 277, 1453-1462, 1997  
A:Title: The complete genome sequence of Escherichia coli K-12.  
A:Reference number: A64720; MUID:97426617; PMID:9278503  
A:Accession: F65011  
A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-80 <BLAT>  
A:Cross-references: GB:AE000326; GB:U00096; NID:g1788718; PIDN:AACT5436.1; PID:g1788720  
A:Experimental source: strain K-12, substrain MG1655

Query Match 31.2%: Score 5; DB 2; Length 80;  
Best Local Similarity 100.0%; Pred. No. 67;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VGYGR 5  
|||||  
Db 75 VGYGR 79

RESULT 22  
J00041  
hypothetical protein, 11K - Escherichia coli insertion sequence IS2  
C:Species: Escherichia coli  
C:Date: 31-Mar-1990 #sequence\_revision 31-Mar-1990 #text\_change 08-Oct-1999  
C:Accession: J00041  
R:Ronecker, H.J.; Rak, B.  
Gene 59, 291-296, 1987  
A:Title: Genetic organization of insertion element IS2 based on a revised nucleotide se  
A:Reference number: A91582; MUID:8813965; PMID:2830172  
A:Accession: J00041  
A:Status: translation not shown  
A:Molecule type: DNA  
A:Residues: 1-97 <RON>  
A:Cross-references: GB:V00279; GB:J01732; GB:M18426; NID:g41525; PIDN:CAA23540.1; PID:g  
A:Experimental source: strain K-12, strain HB101  
C:Genetics:  
A:Mobile element: insertion sequence IS2  
A:Start codon: GTG

Query Match 31.2%: Score 5; DB 2; Length 97;  
Best Local Similarity 100.0%; Pred. No. 80;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VGYGR 5  
|||||  
Db 45 VGYGR 49

RESULT 23  
S64056  
probable membrane protein YGL052w - Yeast (Saccharomyces cerevisiae)  
N:Alternate names: hypothetical protein G3486  
C:Species: Saccharomyces cerevisiae  
C:Date: 17-May-1996 #sequence\_revision 17-May-1996 #text\_change 19-Apr-2002  
R:Feuermann, M.; Potler, S.; Souciet, J.L.  
submitted to the Protein Sequence Database, May 1996  
A:Reference number: S64044  
A:Accession: S64056  
A:Molecule type: DNA  
A:Residues: 1-101 <FEU>  
A:Cross-references: EMBL:Z72574; NID:g1322544; PID:g1322545; GSPDB:GN00007; MIPS:YGL052  
C:Genetics:  
A:Gene: MIPS:YGL052w  
A:Cross-references: SGD:S0003020  
A:Map position: 7L

C:Superfamily: Saccharomyces probable membrane protein YG052w  
C:Keywords: transmembrane protein  
F:13-29/Domain: transmembrane status predicted <TM>  
Query Match 31.2%; Score 5; DB 2; Length 101;  
Best Local Similarity 100.0%; Pred. No. 83;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 11 DPSRK 15  
|||  
Db 86 DPSRK 90  
RESULT 24  
T39293  
hypothetical protein SPBC1105.16c - fission yeast (Schizosaccharomyces pombe)  
C:Species: Schizosaccharomyces pombe  
C>Date: 03-Dec-1999 #sequence\_revision 03-Dec-1999 #text\_change 03-Dec-1999  
C:Accession: T39293  
R:McDougall, R.C.; Rajandream, M.A.; Barrell, B.G.; Ramsperger, U.; Bothe, G.; Pohl, T.  
submitted to the EMBL Data Library, July 1999  
A:Reference number: Z21841  
A:Accession: T39293  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-107 <MCD>  
A:Cross-references: EMBL:AL096851; PIDN:GAB50979.1; GSPDB:GN00067; SPDB:SPBC1105.16c  
A:Experimental source: strain 972h-; cosmid c1105  
C:Genetics:  
A:Gene: SPDB:SPBC1105.16c  
A:Map position: 2  
Query Match 31.2%; Score 5; DB 2; Length 107;  
Best Local Similarity 100.0%; Pred. No. 87;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 12 PSRRN 16  
|||  
Db 79 PSRRN 83  
RESULT 25  
T29627  
hypothetical protein K09E3.2 - Caenorhabditis elegans  
C:Species: Caenorhabditis elegans  
C>Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 04-Mar-2000  
C:Accession: T29627  
R:Johnson, D.; Gattung, S.  
submitted to the EMBL Data Library, November 1995  
A:Description: The sequence of C. elegans cosmid K09E3.  
A:Reference number: Z20655  
A:Accession: T29627  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-109 <JOH>  
A:Cross-references: EMBL:U41033; PIDN:AAA82375.1; CESP:K09E3.2  
C:Genetics:  
A:Gene: CESP:K09E3.2  
C:Superfamily: Caenorhabditis elegans hypothetical protein K09E3.2  
Query Match 31.2%; Score 5; DB 2; Length 109;  
Best Local Similarity 100.0%; Pred. No. 89;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 7 DNDRD 11  
|||  
Db 81 DNDRD 85  
RESULT 26  
C93956  
hypothetical protein SMB21339 (imported) - Sinorhizobium meliloti (strain 1021) magaplas  
C:Species: Sinorhizobium meliloti

C>Date: 24-Aug-2001 #sequence\_revision 24-Aug-2001 #text\_change 30-Sep-2001  
C:Accession: C93956  
R:Finan, T.M.; Weidner, S.; Wong, K.; Buhrmester, J.; Chain, P.; Vorholter, F.J.; Herna  
Proc. Natl. Acad. Sci. U.S.A. 98, 9889-9894, 2001  
A:Title: The complete sequence of the 1,683-kb pSymb megaplasmid from the N2-fixing end  
A:Reference number: A95842; MUID:21396508; PMID:11481431  
A:Accession: C93956  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-110 <KUR>  
A:Cross-references: GB:AL591985; PIDN:CAC49315.1; PID:g15140801; GSPDB:GN00167  
A:Experimental source: strain 1021, megaplasmid pSymb  
R:Galibert, F.; Finan, T.M.; Long, S.R.; Punler, A.; Abola, P.; Ampe, F.; Barloy-Hubler  
pela, D.; Chain, P.; Cowie, A.; Davis, R.W.; Dreano, S.; Federspiel, N.A.; Fisher, R.F.  
L.; Hyman, R.W.; Jones, T.  
A:Authors: Kahn, D.; Kahn, M.L.; Kalman, S.; Keating, D.H.; Kise, E.; Komp, C.; Lelane  
heault, P.; Vandenbol, M.; Vorholter, F.J.; Weidner, S.; Wells, D.H.; Wong, K.; Yen, K.  
A:Title: The composite genome of the legume symbiont Sinorhizobium meliloti.  
A:Reference number: A96039; MUID:21368234; PMID:11474104  
A:Contents: annotation  
C:Genetics:  
A:Gene: SMB21339  
A:Genome: plasmid  
Query Match 31.2%; Score 5; DB 2; Length 110;  
Best Local Similarity 100.0%; Pred. No. 90;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 10 RDPGR 14  
|||  
Db 26 RDPGR 30  
RESULT 27  
PC4215  
phosphatidylethanolamine binding protein Ovd2 - nematode (Onchocerca volvulus) (Fragmen  
C:Species: Onchocerca volvulus  
C>Date: 17-Dec-1996 #sequence\_revision 21-Jan-1997 #text\_change 20-Jun-2000  
C:Accession: PC4215  
R:Ertmann, K.D.; Gallin, M.Y.  
Gene 174, 203-207, 1996  
A:Title: Onchocerca volvulus: Identification of cDNAs encoding a putative phosphatidyl-  
A:Reference number: PC4214; MUID:97045813; PMID:8890725  
A:Accession: PC4215  
A:Status: nucleic acid sequence not shown  
A:Molecule type: mRNA  
A:Residues: 1-114 <ERT>  
A:Cross-references: EMBL:X87990  
C:Comment: This protein is involved in odorant binding and is located in the chemosenso  
C:Superfamily: Caenorhabditis elegans hypothetical protein Y69E1A.5  
Query Match 31.2%; Score 5; DB 2; Length 114;  
Best Local Similarity 100.0%; Pred. No. 93;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 12 PSRRN 16  
|||  
Db 19 PSRRN 23  
RESULT 28  
A27888  
Ig heavy chain V region (H37-84) - mouse  
C:Species: Mus musculus (house mouse)  
C>Date: 15-Dec-1988 #sequence\_revision 15-Dec-1988 #text\_change 16-Aug-1996  
C:Accession: A27888  
R:Gaton, A.J.; Brownlee, G.G.; Staedt, L.M.; Gerhard, W.  
EMBO J. 5, 1577-1587, 1986  
A:Title: Structural and functional implications of a restricted antibody response to a  
A:Reference number: A91043; MUID:86300658; PMID:2427335  
A:Accession: A27888  
A:Molecule type: DNA

A;Residues: 1-121 <CAT>  
A;Experimental source: strain Balb/c  
A;Note: this sequence was determined from the germ-line gene  
C;Comment: This chain was isolated from a hybridoma protein that binds influenza virus h  
C;Superfamily: immunoglobulin V region; immunoglobulin homology  
C;Keywords: heterotetramer; immunoglobulin  
F;15-98/Domain: immunoglobulin homology <IMM>

Query Match 31.2%; Score 5; DB 2; Length 121;  
Best Local Similarity 100.0%; Pred. No. 98;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 7 DNDRD 11  
|||||  
Db 73 DNDRD 77

RESULT 29  
A;Accession: A13574  
A;Species: Brucella melitensis (strain 16M)  
C;Date: 01-Feb-2002 #sequence\_revision 01-Feb-2002 #text\_change 01-Feb-2002  
C;Accession: A13574  
R;DelVecchio, V.G.; Kapural, V.; Redkar, R.U.; Patra, G.; Mujer, C.; Los, T.; Yanova,  
.; Mazur, M.; Goldsman, E.; Selkov, E.; Elizer, P.H.; Hagius, S.; O'Callaghan, D.; Letess  
Proc. Natl. Acad. Sci. U.S.A. 99, 443-448, 2002  
A;Title: The genome sequence of the facultative intracellular pathogen Brucella melitens  
A;Reference number: AD3252; PMID:11756688  
A;Accession: A13574  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-121 <KUR>  
A;Cross-references: GB:AE008918; PIDN:AAL53764.1; PID:G17984692; GSPDB:GN00191  
A;Experimental source: strain 16M  
A;Gene: BMEL10522  
A;Map position: 11

Query Match 31.2%; Score 5; DB 2; Length 121;  
Best Local Similarity 100.0%; Pred. No. 98;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2 GYGRD 6  
|||||  
Db 56 GYGRD 60

RESULT 30  
B69287  
conserved hypothetical protein AF0298 - Archaeoglobus fulgidus  
C;Species: Archaeoglobus fulgidus  
C;Date: 05-Dec-1997 #sequence\_revision 05-Dec-1997 #text\_change 22-Oct-1999  
C;Accession: B69287  
R;Klenk, H.P.; Clayton, R.A.; Tomb, J.F.; White, O.; Nelson, K.E.; Ketchum, K.A.; Dodson  
.; Fleischmann, R.D.; Quackenbush, J.; Lee, N.H.; Sutton, G.G.; Gill, S.; Kirkness, E.F.  
GlobeX, A.; Zhou, L.; Overbeek, R.; Gocayne, J.D.; Weidman, J.F.; McDonald, L.  
Nature 390, 364-370, 1997  
A;Authors: Ullrich, T.; Cotton, M.D.; Spriggs, T.; Attlich, P.; Kaine, B.P.; Sykes, S.  
Smith, H.O.; Weese, C.R.; Venter, J.C.  
A;Title: The complete genome sequence of the hyperthermophilic, sulfate-reducing archaeo  
A;Reference number: A69250; MUID:98049343; PMID:9389475  
A;Accession: B69287  
A;Status: preliminary; nucleic acid sequence not shown; translation not shown  
A;Molecule type: DNA  
A;Residues: 1-125 <KLE>  
A;Cross-references: GB:AE001084; GB:AE000782; NID:G2689407; PIDN:AAB90931.1; PID:G265033

Query Match 31.2%; Score 5; DB 2; Length 125;  
Best Local Similarity 100.0%; Pred. No. 1e+02;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2 GYGRD 6  
|||||

Db 48 GYGRD 52

RESULT 31  
B87434  
hypothetical protein CC1494 [imported] - Caulobacter crescentus  
C;Species: Caulobacter crescentus  
C;Date: 20-Apr-2001 #sequence\_revision 20-Apr-2001 #text\_change 20-Apr-2001  
C;Accession: B87434  
R;Niernern, W.C.; Feldblyum, T.V.; Paulsen, I.T.; Nelson, K.E.; Bisen, J.; Heidelberg, J  
B.; Laub, M.T.; Deboy, R.T.; Dodson, R.J.; Durkin, A.S.; Gwin, M.L.; Haft, D.H.; Kolo  
n, J.; Ermolaeva, M.; White, O.; Salzberg, S.L.; Shapiro, L.; Venter, J.C.; Fraser, C.M  
Proc. Natl. Acad. Sci. U.S.A. 98, 4136-4141, 2001  
A;Title: Complete Genome Sequence of Caulobacter crescentus.  
A;Reference number: A87249; MUID:21173698; PMID:11259647  
A;Accession: B87434  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-127 <STO>  
A;Cross-references: GB:AE005673; NID:G13422869; PIDN:AAK23473.1; GSPDB:GN00148  
C;Gene: CC1494

Query Match 31.2%; Score 5; DB 2; Length 127;  
Best Local Similarity 100.0%; Pred. No. 1e+02;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 12 PSRKN 16  
|||||  
Db 114 PSRKN 118

RESULT 32  
S76955  
hypothetical protein - Synechocystis sp. (strain PCC 6803)  
C;Species: Synechocystis sp.  
A;Variety: PCC 6803  
C;Date: 25-Apr-1997 #sequence\_revision 25-Apr-1997 #text\_change 08-Oct-1999  
C;Accession: S76955  
R;Kaneko, T.; Sato, S.; Kozaki, H.; Tanaka, A.; Asamizu, E.; Nakamura, Y.; Miyajima, N.  
c, K.; Okumura, S.; Shingo, S.; Takeuchi, C.; Wada, T.; Watanabe, A.; Yamada, M.; Yasud  
DNA Res. 3, 109-136, 1996  
A;Title: Sequence analysis of the genome of the unicellular cyanobacterium Synechocysti  
s.  
A;Reference number: S74322; MUID:97061201; PMID:8905231  
A;Accession: S76955  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-128 <KAN>  
A;Cross-references: EMBL:D90917; GB:AB001339; NID:G1653836; PIDN:BA18867.1; PID:dl0196  
A;Note: the nucleotide sequence was submitted to the EMBL Data Library, June 1996

Query Match 31.2%; Score 5; DB 2; Length 128;  
Best Local Similarity 100.0%; Pred. No. 1e+02;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 11 DPSRK 15  
|||||  
Db 24 DPSRK 28

RESULT 33  
A75199  
hypothetical protein PA82271 - Pyrococcus abyssi (strain Orsay)  
C;Species: Pyrococcus abyssi  
C;Date: 20-Aug-1999 #sequence\_revision 20-Aug-1999 #text\_change 20-Aug-1999  
C;Accession: A75199  
R;Anonymous, GenomeScope  
submitted to the EMBL Data Library, July 1999  
A;Description: Pyrococcus abyssi genome sequence: insights into archaeal chromosome str  
A;Reference number: A75001  
A;Accession: A75199  
A;Status: preliminary

A;Molecule type: DNA  
A;Residues: 1-128 <KAM>  
A;Cross-references: GB:AJ248283; GB:AL096836; NID:G5457433; PIDN:CAB49040.1; PID:G151493  
A;Experimental source: strain Oresay  
C;Genetics:  
A;Gene: PAB2271

Query Match 31.2%; Score 5; DB 2; Length 128;  
Best Local Similarity 100.0%; Pred. No. 1e+02;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 GYGRD 6  
|||||  
DB 51 GYGRD 55

RESULT 34  
A96706  
unknown protein, 64033-64434 [imported] - Arabidopsis thaliana  
C;Species: Arabidopsis thaliana (mouse-ear cress)  
C;Date: 02-Mar-2001 #sequence\_revision 02-Mar-2001 #text\_change 31-Mar-2001  
C;Accession: A96706  
R;Theologos, A.; Ecker, J.R.; Palm, C.J.; Federpiel, N.A.; Kaul, S.; White, O.; Alonso, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Creasy, T.H.; Dewar, K.; ansen, N.F.; Hughes, B.; Huizart, L.  
A;Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Maiti, R.; Marzilli, Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon, A.;Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon, ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.  
A;Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.  
A;Reference number: A86141; PMID:21016719; PMID:11130712  
A;Accession: A96706  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-133 <STO>  
A;Cross-references: GB:AB05173; NID:G675727; PIDN:AF26488.1; GSPDB:GN00141  
C;Genetics:  
A;Gene: T22E19.12  
A;Map position: 1

Query Match 31.2%; Score 5; DB 2; Length 133;  
Best Local Similarity 100.0%; Pred. No. 1.1e+02;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 7 DNRD 11  
|||||  
DB 2 DNRD 6

RESULT 35  
PC4214  
phosphatidylethanolamine binding protein Ovd3 - nematode (Onchocerca volvulus) (fragment  
C;Species: Onchocerca volvulus  
C;Date: 17-Dec-1996 #sequence\_revision 21-Jan-1997 #text\_change 20-Jun-2000  
C;Accession: PC4214  
R;Bertmann, K.D.; Gallin, M.Y.  
Gene 174, 203-207, 1996  
A;Title: Onchocerca voluulus: Identification of cDNAs encoding a putative phosphatidyl-e  
A;Reference number: PC4214; MUID:97045813; PMID:8890735  
A;Accession: PC4214  
A;Molecule type: mRNA  
A;Residues: 1-134 <ERT>  
A;Cross-references: EMBL:X87989; NID:G1143530; PID:G1143531  
C;Genetics:  
A;Intons: 34/3; 52/2; 81/1; 129/3  
C;Superfamily: Caenorhabditis elegans hypothetical protein Y69E1A.5

Query Match 31.2%; Score 5; DB 2; Length 134;  
Best Local Similarity 100.0%; Pred. No. 1.1e+02;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 12 PSRRN 16  
|||||  
DB 39 PSRRN 43

RESULT 36  
AG3272  
probable lactoylgutathione lyase (EC 4.4.1.5) [imported] - Brucella melitensis (strain  
C;Species: Brucella melitensis  
C;Date: 01-Feb-2002 #sequence\_revision 01-Feb-2002 #text\_change 22-Mar-2002  
C;Accession: AG3272  
R;Deivechho, V.G.; Kapatal, V.; Redkar, R.J.; Patra, G.; Mujer, C.; Ios, T.; Ivanova, .; Mazur, M.; Goldsman, E.; Selkov, E.; Elzer, P.H.; Hagius, S.; O'Callaghan, D.; Leter Proc. Natl. Acad. Sci. U.S.A. 99, 443-448, 2002  
A;Title: The genome sequence of the facultative intracellular pathogen Brucella melite  
A;Reference number: AD3252; PMID:11756688  
A;Accession: AG3272  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-134 <KUR>  
A;Cross-references: GB:AF008917; PIDN:AL51346.1; PID:G17982044; GSPDB:GN00190  
A;Experimental source: strain 16M  
C;Genetics:  
A;Gene: BME10164  
A;Map position: 1  
C;Keywords: carbon-sulfur lyase

Query Match 31.2%; Score 5; DB 2; Length 134;  
Best Local Similarity 100.0%; Pred. No. 1.1e+02;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VGYGR 5  
|||||  
DB 49 VGYGR 53

RESULT 37  
AF0812  
probable acetyltransferase STY2686 [imported] - Salmonella enterica subsp. enterica ser  
C;Species: Salmonella enterica subsp. enterica serovar Typh  
A;Note: this species has also been called Salmonella typh  
C;Date: 09-Nov-2001 #sequence\_revision 09-Nov-2001 #text\_change 09-Nov-2001  
C;Accession: AF0812  
R;Parkhill, J.; Dougan, G.; James, K.D.; Thomson, N.R.; Pickard, D.; Wain, J.; Churcher th, T.; Connerton, P.; Cronin, A.; Davis, P.; Davies, R.M.; Dowd, L.; White, N.; Farrar . S.; Moule, S.; O'Garra, P.  
Nature 413, 848-852, 2001  
A;Authors: Parry, C.; Quail, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K. A;Title: Complete genome sequence of a multiple drug resistant Salmonella enterica sero  
A;Reference number: AB0502; PMID:11677608  
A;Accession: AF0812  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-141 <PAR>  
A;Cross-references: GB:AL51382; PIDN:CAD07680.1; PID:G16503666; GSPDB:GN00176  
C;Genetics:  
A;Gene: STY2686

Query Match 31.2%; Score 5; DB 2; Length 141;  
Best Local Similarity 100.0%; Pred. No. 1.1e+02;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 5 RDDND 9  
|||||  
DB 108 RDDND 112

RESULT 38  
A97008  
transcription regulator, Warr/Emr family [imported] - Clostridium acetobutylicum  
C;Species: Clostridium acetobutylicum  
C;Date: 14-Sep-2001 #sequence\_revision 14-Sep-2001 #text\_change 14-Sep-2001  
C;Accession: A97008

R;Nolling, J.; Breton, G.; Omeichenko, M.V.; Markarova, K.S.; Zeng, Q.; Gibson, R.; Lee, J.; Daly, M.J.; Bennett, G.N.; Koonin, E.V.; Smith, D.R.  
 J. Bacteriol. 183, 4823-4838, 2001  
 A;Title: Genome Sequence and Comparative Analysis of the Solvent-Producing Bacterium ClC  
 A;Reference number: A96900; MUID:21359325; PMID:21359325  
 A;Accession: A97008

A;Status: preliminary  
 A;Molecule type: DNA

A;Residues: 1-151 <KUR>

A;Cross-references: GB:AE001437; PIDN:AAK78952.1; PID:G15023772; GSPDB:GN00168  
 A;Experimental source: Clostridium acetobutylicum ATCC824

C;Genetics:  
 A;Gene: CAC0876

Query Match 31.2%; Score 5; DB 2; Length 151;  
 Best Local Similarity 100.0%; Pred. No. 1.2e+02;

Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 6 DDNDR 10  
 |||||

Db 88 DDNDR 92

#### RESULT 39

PC4216  
 phosphatidylethanolamine binding protein Ovd1 - nematode (Onchocerca volvulus) (fragment  
 C;Species: Onchocerca volvulus

C;Date: 17-Dec-1996 #sequence\_revision 21-Jan-1997 #text\_change 20-Jun-2000  
 C;Accession: PC4216

R;Bittmann, K.D.; Gallin, M.Y.  
 Gene 174, 203-207, 1996

A;Title: Onchocerca volvulus: Identification of cDNAs encoding a putative phosphatidyl-  
 A;Reference number: PC4214; MUID:97045813; PMID:8890735  
 A;Accession: PC4216

A;Status: nucleic acid sequence not shown

A;Molecule type: mRNA

A;Residues: 1-152 <ERT>

A;Cross-references: EMBL:X87991  
 C;Comment: This protein is involved in odorant binding and is located in the chemosensor  
 C;Superfamily: Caenorhabditis elegans hypothetical protein Y69E1A.5

Query Match 31.2%; Score 5; DB 2; Length 152;  
 Best Local Similarity 100.0%; Pred. No. 1.2e+02;  
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 12 PSRKN 16  
 |||||

Db 57 PSRKN 61

#### RESULT 40

A44947  
 1cK antigen precursor - nematode (Onchocerca volvulus)

C;Species: Onchocerca volvulus  
 C;Date: 14-May-1993 #sequence\_revision 14-May-1993 #text\_change 20-Jun-2000  
 C;Accession: A44947

R;Lobos, E.; Altmann, M.; Mengod, G.; Weiss, N.; Rudin, W.; Karam, M.  
 Mol. Biochem. Parasitol. 39, 135-146, 1990

A;Title: Identification of an Onchocerca volvulus cDNA encoding a low-molecular-weight  
 A;Reference number: A44947; MUID:90158691; PMID:1689459  
 A;Accession: A44947

A;Status: preliminary

A;Molecule type: mRNA

A;Residues: 1-152 <LOB>

A;Cross-references: GB:M27807; NID:G159872; PID:G159873  
 C;Superfamily: Caenorhabditis elegans hypothetical protein Y69E1A.5

Query Match 31.2%; Score 5; DB 2; Length 152;  
 Best Local Similarity 100.0%; Pred. No. 1.2e+02;  
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 12 PSRKN 16  
 |||||

Db 102 PSRKN 106

Search completed: April 9, 2003, 13:16:13  
 Job time : 21.8197 secs

Result	Score	Query	Match	length	DB	ID	Description
1	552	100.0	1477	9	US-10-020-441-1		Sequence 1, Appl1
2	95.5	17.3	897	10	US-09-923-779-146		Sequence 146, Appl1
3	95.5	17.3	966	10	US-09-923-237-236		Sequence 236, Appl1
4	93	16.8	1081	10	US-09-954-456-2126		Sequence 2126, Appl1

C	5	87.5	15.9	937	10	US-09-920-345-22	Sequence 22, App
6	84	15.2	3987	10	US-09-888-615-38	Sequence 38, App	
7	83.5	15.1	1110	9	US-10-041-006A-1	Sequence 1, App	
8	83.5	15.1	1110	12	US-10-040-655-1	Sequence 1, App	
9	83.5	15.1	1129	9	US-10-028-072-22	Sequence 22, App	
10	83.5	15.1	1129	9	US-10-121-049-221	Sequence 22, App	
11	83.5	15.1	1129	9	US-10-123-904-221	Sequence 22, App	
12	83.5	15.1	1129	9	US-10-140-470-221	Sequence 22, App	
13	83.5	15.1	1129	9	US-10-175-746-221	Sequence 22, App	
14	83.5	15.1	1129	9	US-10-176-918-221	Sequence 22, App	
15	83.5	15.1	1129	9	US-10-176-921-221	Sequence 22, App	
16	83.5	15.1	1129	9	US-10-137-865-221	Sequence 22, App	
17	83.5	15.1	1129	9	US-10-140-474-221	Sequence 22, App	
18	83.5	15.1	1129	9	US-10-142-431-221	Sequence 22, App	
19	83.5	15.1	1129	9	US-10-143-114-221	Sequence 22, App	
20	83.5	15.1	1129	9	US-10-140-002-221	Sequence 22, App	
21	83.5	15.1	1129	9	US-10-144-419-221	Sequence 22, App	
22	83.5	15.1	1129	9	US-10-123-262-221	Sequence 22, App	
23	83.5	15.1	1129	9	US-10-142-423-221	Sequence 22, App	
24	83.5	15.1	1129	9	US-10-121-050-221	Sequence 22, App	
25	83.5	15.1	1129	9	US-10-141-755-221	Sequence 22, App	
26	83.5	15.1	1129	9	US-10-143-032-221	Sequence 22, App	
27	83.5	15.1	1130	9	US-10-041-006A-8	Sequence 8, App	
28	83.5	15.1	1130	12	US-10-040-655-8	Sequence 8, App	
29	83	15.0	7350	10	US-09-768-826-16	Sequence 16, App	
30	82.5	14.9	726	9	US-10-092-000A-1	Sequence 1, App	
31	82.5	14.9	3147	9	US-10-092-700A-1	Sequence 1, App	
32	82.5	14.9	3147	9	US-10-092-700A-3	Sequence 3, App	
33	80.5	14.6	2756	10	US-09-925-301-351	Sequence 35, App	
34	80	14.5	1130	12	US-10-041-400A-8	Sequence 8, App	
35	80	14.5	1130	12	US-10-041-264A-8	Sequence 8, App	
36	80	14.5	1130	12	US-10-042-091A-8	Sequence 8, App	
37	80	14.5	1613	12	US-10-041-000A-1	Sequence 1, App	
38	80	14.5	1613	12	US-10-041-264A-1	Sequence 1, App	
39	80	14.5	1613	12	US-10-042-091A-1	Sequence 1, App	
40	80	14.5	1836	10	US-09-832-197-2	Sequence 2, App	
41	80	14.5	1887	10	US-09-888-615-27	Sequence 27, App	
42	80	14.5	2365	9	US-09-978-295A-131	Sequence 131, App	
43	80	14.5	2365	9	US-09-978-697-131	Sequence 131, App	
44	80	14.5	2365	9	US-09-978-182A-131	Sequence 131, App	
45	80	14.5	2365	9	US-09-999-832A-131	Sequence 131, App	

## ALIGNMENTS

```

1 RESULT 1
2 US-10-020-441-1
3 ; Sequence 1, Application US/10020441
4 ; Publication No. US20020182224A1
5 ; GENERAL INFORMATION:
6 ; APPLICANT: Doenhoff, Michael
7 ; APPLICANT: Sayers, Jon
8 ; TITLE OF INVENTION: SCHISTOSOMIASIS VACCINE
9 ; FILE REFERENCE: Doenhoff
10 ; CURRENT APPLICATION NUMBER: US/10/020,441
11 ; CURRENT FILING DATE: 2001-12-18
12 ; PRIOR APPLICATION NUMBER: US/09/413,810
13 ; PRIOR FILING DATE: PRIOR FILING DATE: 1999-10-07
14 ; NUMBER OF SEQ. ID NOS: 3
15 ; SOFTWARE: PatentIn Ver. 2.0
16 SEQ. ID NO. 1
17 LENGTH: 1477
18 ; TYPE: DNA
19 ; ORGANISM: Schistosoma mansoni
20 US-10-020-441-1

```

Alignment Scores:	
Pred. No.:	4,656-67
Score:	552.00
Percent Similarity:	100.00%
Best Local Similarity:	100.00%
Query Match:	100.00%
BB:	9
Gaps:	0
Length:	147
Matches:	106
Conservative:	0
Mismatches:	0
Indels:	0
Gaps:	0

US-10-020-441-2 (1-106) x US-10-020-441-1 (1-1477)

Qy 1 ValSerPheLeuThrLeuArgAsnGlyAspGlnGlnGlyIleHisGlnProSerGly 20  
 Db 310 GTTATATTTCTACACGTGAGATGGCGACCAAGGATCATCATCCACCGTCTGGA 369  
 Qy 21 ValIleValAlaProGlyTyrMetProSerCysMetSerAlaArgGlnArgProIle 40  
 Db 370 GTTAAGTGTGCACACGATACATGCCCTTGTATGTCCGACGACAGAGAGACCAATC 429  
 Qy 41 AlaGlnThrLeuSerGlyPheAspIleAlaIleValMetLeuAlaGlnMetValAsnLeu 60  
 Db 430 GCACAGACACACTGATGATTCATATTCATATGCAATGTCTGGCTCAATGTCAACTTA 489  
 Qy 61 GlnSerGlyIleArgValIleSerLeuProGlnProSerAspIleProProGlyThr 80  
 Db 490 CAGAGTGAATCAGATGATCAGTCTGCCACAGCCATCGGATATCCCGCACCTGGAACT 549  
 Qy 81 GlyValPheIleValGlyTyrGlyArgAspAspAsnAspArgAspProSerArgIysAsn 100  
 Db 550 GGTGTTTTCATTTGTTGTTATGGAAGGATGATACGACCGTGATCCGTACGTAAGAT 609  
 Qy 101 GlyGlyIleLeuIlysIys 106  
 Db 610 GGTGGAATATGAGAGAA 627

# RESULT 2

US-09-923-779-146  
 ; Sequence 146, Application US/09923779  
 ; Patent No. US20020076721A1  
 ; GENERAL INFORMATION:

APPLICANT: Pyle, Ruth A.  
 APPLICANT: Xu, Jiangchun  
 APPLICANT: Kalos, Michael D.  
 TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY  
 ; TITLE OF INVENTION: AND DIAGNOSIS OF PANCREATIC CANCER  
 ; FILE REFERENCE: 210121.553  
 ; CURRENT APPLICATION NUMBER: US/09/923,779  
 ; CURRENT FILING DATE: 2001-08-06  
 ; NUMBER OF SEQ ID NOS: 155  
 ; SOFTWARE: FastSeq for Windows Version 4.0  
 ; SEQ ID NO 146  
 ; LENGTH: 897  
 ; TYPE: DNA  
 ; ORGANISM: Homo sapiens  
 ; US-09-923-779-146

# Alignment Scores:

Pred. No.:	Score:	Length:	Matches:	Conservative:	Mismatches:	Indels:	Gaps:
0.000485	95.50	897	25	16	34	11	1
Percent Similarity:	47.67%						
Best Local Similarity:	29.07%						
Query Match:	17.30%						
DB:	10						

US-10-020-441-2 (1-106) x US-09-923-779-146 (1-897)

Qy 4 LeuThrLeuArgAsnGlyAspGlnGlnGlyIleHisGlnProSerGlyValIysVal 23  
 Db 289 CTTGCTGTGAAGAGGAGGCCCCGACGAGTGATCCCATCAACTCTGAGAGCTGTTTG 348  
 Qy 24 AlaProGlyTyrMetProSerCysMetSerAlaArgGlnArgProIleAlaGlnThr 43  
 Db 349 CATCCACTCTGGAACCGCTGCTGTGTGGCTGT----- 381  
 Qy 44 LeuSerGlyPheAspIleAlaIleValMetLeuAlaGlnMetValAsnLeuGlnSerGly 63  
 Db 382 -----GGCAATACATCCGCCCTCATCAAGCTCTCAACGAGCCGCCAGCTGGAGATGCC 435  
 Qy 64 IleArgValIleSerLeuProGlnProSerAspIleProProGlyThrGlyValPhe 83  
 Db 436 GTTCACACTGCGCTCACTCCCTCCCGCTGTGATCATCTTCCCAACAAGACACCCCTGTAC 495

Qy 84 IleValGlyTyrGlyArg 89  
 Db 496 ATCACCAGGCTGGGGCCGT 513

# RESULT 3

US-09-925-297-236  
 ; Sequence 236, Application US/09925297  
 ; Patent No. US20020081659A1  
 ; GENERAL INFORMATION:

APPLICANT: Rosen et al.  
 TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies  
 ; FILE REFERENCE: PA105  
 ; CURRENT APPLICATION NUMBER: US/09/925,297  
 ; CURRENT FILING DATE: 2001-08-10  
 ; PRIOR APPLICATION NUMBER: PCT/US00/05989  
 ; PRIOR FILING DATE: 2000-03-08  
 ; PRIOR APPLICATION NUMBER: 60/124,270  
 ; PRIOR FILING DATE: 1999-03-12  
 ; NUMBER OF SEQ ID NOS: 928  
 ; SOFTWARE: PatentIn Ver. 2.0  
 ; SEQ ID NO 236  
 ; LENGTH: 966  
 ; TYPE: DNA  
 ; ORGANISM: Homo sapiens  
 ; FEATURE:  
 ; NAME/KEY: misc feature  
 ; LOCATION: (896)  
 ; OTHER INFORMATION: n equals a,t,g, or c  
 ; NAME/KEY: misc feature  
 ; LOCATION: (956)  
 ; OTHER INFORMATION: n equals a,t,g, or c  
 ; US-09-925-297-236

# Alignment Scores:

Pred. No.:	Score:	Length:	Matches:	Conservative:	Mismatches:	Indels:	Gaps:
0.000535	95.50	966	25	16	34	11	1
Percent Similarity:	47.67%						
Best Local Similarity:	29.07%						
Query Match:	17.30%						
DB:	10						

US-10-020-441-2 (1-106) x US-09-925-297-236 (1-966)

Qy 4 LeuThrLeuArgAsnGlyAspGlnGlnGlyIleHisGlnProSerGlyValIysVal 23  
 Db 278 CTTGCTGTGAAGAGGAGGCCCCGACGAGTGATCCCATCAACTCTGAGAGCTGTTTG 337  
 Qy 24 AlaProGlyTyrMetProSerCysMetSerAlaArgGlnArgProIleAlaGlnThr 43  
 Db 338 CATCCACTCTGGAACCGCTGCTGTGTGGCTGT----- 370  
 Qy 44 LeuSerGlyPheAspIleAlaIleValMetLeuAlaGlnMetValAsnLeuGlnSerGly 63  
 Db 371 -----GGCAATGACATCCGCCCTCATCAAGCTCTCAACGAGCCGCCAGCTGGAGATGCC 424  
 Qy 64 IleArgValIleSerLeuProGlnProSerAspIleProProGlyThrGlyValPhe 83  
 Db 425 GTTCACACTGCGCTCACTCCCTCCCGCTGTGATCATCTTCCCAACAAGACACCCCTGTAC 484  
 Qy 84 IleValGlyTyrGlyArg 89  
 Db 485 ATCACCAGGCTGGGGCCGT 502

# RESULT 4

US-09-954-456-2126  
 ; Sequence 2126, Application US/09954456  
 ; Patent No. US20020115057A1  
 ; GENERAL INFORMATION:

APPLICANT: Young, Paul  
 TITLE OF INVENTION: Process for Identifying Anti-Cancer Therapeutic Agents Using Canc  
 ; TITLE OF INVENTION: Sets  
 ; FILE REFERENCE: 689290-76







PRIOR FILING DATE:	1998-02-04
PRIOR APPLICATION NUMBER:	60/074086
PRIOR FILING DATE:	1998-02-09
PRIOR APPLICATION NUMBER:	60/074092
PRIOR FILING DATE:	1998-02-09
PRIOR APPLICATION NUMBER:	60/077791
PRIOR FILING DATE:	1998-03-12
PRIOR APPLICATION NUMBER:	60/078910
PRIOR FILING DATE:	1998-03-20
PRIOR APPLICATION NUMBER:	60/079294
PRIOR FILING DATE:	1998-03-25
PRIOR APPLICATION NUMBER:	60/079666
PRIOR FILING DATE:	1998-02-27
PRIOR APPLICATION NUMBER:	60/079722
PRIOR FILING DATE:	1998-03-27
PRIOR APPLICATION NUMBER:	60/080165
PRIOR FILING DATE:	1998-03-31
PRIOR APPLICATION NUMBER:	60/081200
PRIOR FILING DATE:	1998-04-09
PRIOR APPLICATION NUMBER:	60/081222
PRIOR FILING DATE:	1998-04-15
PRIOR APPLICATION NUMBER:	60/081818
PRIOR FILING DATE:	1998-04-15
PRIOR APPLICATION NUMBER:	60/082999
PRIOR FILING DATE:	1998-04-29
PRIOR APPLICATION NUMBER:	60/084600
PRIOR FILING DATE:	1998-05-07
PRIOR APPLICATION NUMBER:	60/084622
PRIOR FILING DATE:	1998-05-07
PRIOR APPLICATION NUMBER:	60/084637
PRIOR FILING DATE:	1998-05-07
PRIOR APPLICATION NUMBER:	60/085148
PRIOR FILING DATE:	1998-05-12
PRIOR APPLICATION NUMBER:	60/085322
PRIOR FILING DATE:	1998-05-13
PRIOR APPLICATION NUMBER:	60/085338
PRIOR FILING DATE:	1998-05-13
PRIOR APPLICATION NUMBER:	60/085339
PRIOR FILING DATE:	1998-05-13
PRIOR APPLICATION NUMBER:	60/085579
PRIOR FILING DATE:	1998-05-15
PRIOR APPLICATION NUMBER:	60/085697
PRIOR FILING DATE:	1998-05-15
PRIOR APPLICATION NUMBER:	60/085704
PRIOR FILING DATE:	1998-05-15
PRIOR APPLICATION NUMBER:	60/086414
PRIOR FILING DATE:	1998-05-22
PRIOR APPLICATION NUMBER:	60/086430
PRIOR FILING DATE:	1998-05-22
PRIOR APPLICATION NUMBER:	60/087106
PRIOR FILING DATE:	1998-05-28
PRIOR APPLICATION NUMBER:	60/088028
PRIOR FILING DATE:	1998-06-04
PRIOR APPLICATION NUMBER:	60/0880730
PRIOR FILING DATE:	1998-06-10
PRIOR APPLICATION NUMBER:	60/0880741
PRIOR FILING DATE:	1998-06-10
PRIOR APPLICATION NUMBER:	60/088810
PRIOR FILING DATE:	1998-06-10
PRIOR APPLICATION NUMBER:	60/088858
PRIOR FILING DATE:	1998-06-11
PRIOR APPLICATION NUMBER:	60/089532
PRIOR FILING DATE:	1998-06-17
PRIOR APPLICATION NUMBER:	60/089599
PRIOR FILING DATE:	1998-06-17

```

; PRIOR APPLICATION NUMBER: 60/089907
; PRIOR FILING DATE: 1998-06-18
; PRIOR APPLICATION NUMBER: 60/089947
; PRIOR FILING DATE: 1998-06-19
; PRIOR APPLICATION NUMBER: 60/090349
; PRIOR FILING DATE: 1998-06-23
; PRIOR APPLICATION NUMBER: 60/090429
; PRIOR FILING DATE: 1998-06-24
; PRIOR APPLICATION NUMBER: 60/090445
; PRIOR FILING DATE: 1998-06-24
; PRIOR APPLICATION NUMBER: 60/090538
; PRIOR FILING DATE: 1998-06-24
; PRIOR APPLICATION NUMBER: 60/090863
; PRIOR FILING DATE: 1998-06-26
; PRIOR APPLICATION NUMBER: 60/091360
; PRIOR FILING DATE: 1998-07-01
; PRIOR APPLICATION NUMBER: 60/091519
; PRIOR FILING DATE: 1998-07-02
; PRIOR APPLICATION NUMBER: 60/091982
; PRIOR FILING DATE: 1998-07-07
```

## Alignment Scores:

Pred. No.:	0.0303	Length:	1129
Score:	83.50	Matches:	32
Percent Similarity:	46.46%	Conservative:	14
Best Local Similarity:	32.32%	Mismatches:	44
Query Match:	15.13%	Indels:	10
DB:	9	Gaps:	4

US-10-020-441-2 (1-106) x US-10-028-072-221 (1-1129)

```

Qy 7 ArgasnglyAspgInglnglyIleHsiSgInProserglyVallys-----Valala 24
Db 267 CGCAACACCTCTGAGACGTCCTCTGACCAAGTCCTCTGGGGGCAAGCAGCTAGTGAG 326
Qy 25 ProglyTyrmelProserCyMeSerAlaArgGlnArg-----Proile 40
Db 327 CCGGGA-----CCACACGCTATGTATGCCGGGTGAGGAGGAGCAACCCCTG 380
Qy 41 AlaGln---ThrluSerglyPheAspIleAlaIleValMetLeuAlaGlnMetValasn 59
Db 381 TACACGGGACCGCCTCCAGCGCTGACGTGCCCTGTGAGCTGAGGACCAACGCCCC 440
Qy 60 LeuGlnSerglyIleArgValIleSerLeuProGlnProSerAspIleProProgly 79
Db 441 TTCACCAATTACATCTCCCGCTGCTGCTGACCCCTCGGTGATCTTTGAGACGGGC 500
Qy 80 ThrGlyValPheIleValGlyTyrglyArgAspAspAsnAspArgAspProserArg 98
Db 501 ATGAACCTGCTGGGTCACTGCTGGGCGACCCCAAGAGGA--AGACCTCTGCCCCG 556
```

## RESULT 10

```

US-10-121-049-221
; Sequence 221, Application US/10121049
; Publication No. US2003002239A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Beresini, Maureen
; APPLICANT: Deforge, Laura
; APPLICANT: Desnoyers, Luc
; APPLICANT: Filvaroff, Ellen
; APPLICANT: Gao, Wei-Qiang
; APPLICANT: Gerritsen, Mary E.
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Sherwood, Steven
; APPLICANT: Smith, Victoria
; APPLICANT: Stewart, Timothy A.
; APPLICANT: Tumas, Daniel
; APPLICANT: Watanabe, Colin K
; APPLICANT: Wood, William
; APPLICANT: Zhang, Zemin
```

```

; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; FILE OF INVENTION: ACIDS ENCODING THE SAME
; FILE REFERENCE: P3330R1C17
; CURRENT APPLICATION NUMBER: US/10/121,049
; CURRENT FILING DATE: 2002-04-12
; Prior Application removed - See File Wrapper or Palm
; NUMBER OF SEQ ID NOS: 550
; SEQ ID NO 221
; LENGTH: 1129
; TYPE: DNA
; ORGANISM: Homo Sapien
US-10-121-049-221
```

## Alignment Scores:

Pred. No.:	0.0303	Length:	1129
Score:	83.50	Matches:	32
Percent Similarity:	46.46%	Conservative:	14
Best Local Similarity:	32.32%	Mismatches:	44
Query Match:	15.13%	Indels:	10
DB:	9	Gaps:	4

US-10-020-441-2 (1-106) x US-10-121-049-221 (1-1129)

```

Qy 7 ArgasnglyAspgInglnglyIleHsiSgInProserglyVallys-----Valala 24
Db 267 CGCAACACCTCTGAGACGTCCTCTGACCAAGTCCTCTGGGGGCAAGCAGCTAGTGAG 326
Qy 25 ProglyTyrmelProserCyMeSerAlaArgGlnArg-----Proile 40
Db 327 CCGGGA-----CCACACGCTATGTATGCCGGGTGAGGAGGAGCAACCCCTG 380
Qy 41 AlaGln---ThrluSerglyPheAspIleAlaIleValMetLeuAlaGlnMetValasn 59
Db 381 TACACGGGACCGCCTCCAGCGCTGACGTGCCCTGTGAGCTGAGGACCAACGCCCC 440
Qy 60 LeuGlnSerglyIleArgValIleSerLeuProGlnProSerAspIleProProgly 79
Db 441 TTCACCAATTACATCTCCCGCTGCTGCTGACCCCTCGGTGATCTTTGAGACGGGC 500
Qy 80 ThrGlyValPheIleValGlyTyrglyArgAspAspAsnAspArgAspProserArg 98
Db 501 ATGAACCTGCTGGGTCACTGCTGGGCGACCCCAAGAGGA--AGACCTCTGCCCCG 556
```

## RESULT 11

```

US-10-123-904-221
; Sequence 221, Application US/10123904
; Publication No. US2003002239A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Beresini, Maureen
; APPLICANT: Deforge, Laura
; APPLICANT: Desnoyers, Luc
; APPLICANT: Filvaroff, Ellen
; APPLICANT: Gao, Wei-Qiang
; APPLICANT: Gerritsen, Mary E.
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Sherwood, Steven
; APPLICANT: Smith, Victoria
; APPLICANT: Stewart, Timothy A.
; APPLICANT: Tumas, Daniel
; APPLICANT: Watanabe, Colin K
; APPLICANT: Wood, William
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; FILE OF INVENTION: ACIDS ENCODING THE SAME
; FILE REFERENCE: P3330R1C54
; CURRENT APPLICATION NUMBER: US/10/123,904
; CURRENT FILING DATE: 2002-04-16
; Prior Application removed - See File Wrapper or Palm
; NUMBER OF SEQ ID NOS: 550
; SEQ ID NO 221
```

LENGTH: 1129  
TYPE: DNA  
ORGANISM: Homo Sapien  
US-10-123-904-221

## Alignment Scores:

Pred. No.: 0.0303 Length: 1129  
Score: 83.50 Matches: 32  
Percent Similarity: 46.46% Conservative: 14  
Best Local Similarity: 32.32% Mismatches: 44  
Query Match: 15.13% Indels: 10  
DB: 9 Gaps: 4

US-10-020-441-2 (1-106) x US-10-123-904-221 (1-1129)

Qy 7 ArgAenGIYAepGInGInGIYlIehIshIsgInProSeRGIYValIys-----ValAla 24  
Db 267 CGCAACACCTCTGAGACGCTCCCTGTACAGAGTCTGCTGGGGGCAAGCAGCTAGTGAG 326  
Qy 25 ProGIYTYMeTProSeRGIYMeTSeRAlAArgInAArg-----ProIle 40  
Db 327 CCGGGA-----CCACAGCTATGTATGCCCCGGGTGAGGCAAGTGAAGCAACCCCTG 380  
Qy 41 AlAgiN---ThrLeuSeRGIYpHeaSpIleAlIeValIeMeTleuAlaGIImeTValAsn 59  
Db 381 TACCAAGGACGAGCCTCCAGCGCTGACGTGGCCCTGTGAGAGTGAAGCAGCAGGCC 440  
Qy 60 LeuGIInSeRGIYlIeArgValIleSeRleuProGIInProSeRAspIleProProGIY 79  
Db 441 TTCACCACTTACATCTCTCCCTCCGCTGCTGACCTGACCTCGTATCTTTGAGACGGGC 500  
Qy 80 ThrGIYValIpHeIleValIGlYTYRGIYAArgAspAspAspAspProSeRArg 98  
Db 501 ATGAACCTGCTGGTCACTGCTGGGGCAGCCCAAGTGAAGA-AGACCTCTGCCCCA 556

## RESULT 12

US-10-140-470-221  
Sequence 221, Application US/10140470  
Publication No. US2003002231A1

## GENERAL INFORMATION:

APPLICANT: Baker, Kevin P.  
APPLICANT: Beresini, Maureen  
APPLICANT: Deforge, Laura  
APPLICANT: Desnoyers, Luc  
APPLICANT: Filvaroff, Ellen  
APPLICANT: Gao, Wei-Qiang  
APPLICANT: Gerritsen, Mary E.  
APPLICANT: Goddard, Audrey  
APPLICANT: Godowski, Paul J.  
APPLICANT: Gurney, Austin L.  
APPLICANT: Sherwood, Steven  
APPLICANT: Smith, Victoria  
APPLICANT: Stewart, Timothy A.  
APPLICANT: Tumas, Daniel  
APPLICANT: Watanabe, Colin K  
APPLICANT: Wood, William  
APPLICANT: Zhang, Zemin  
TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC  
FILE REFERENCE: P330R1C160  
CURRENT APPLICATION NUMBER: US/10/140,470  
CURRENT FILING DATE: 2002-05-06  
Prior Application removed - See Palm or File Wrapper  
NUMBER OF SEQ ID NOS: 550  
SEQ ID NO 221  
LENGTH: 1129  
TYPE: DNA  
ORGANISM: Homo Sapien  
US-10-140-470-221

Alignment Scores: 0.0303 Length: 1129  
Pred. No.: 83.50 Matches: 32  
Score: 83.50

Percent Similarity: 46.46% Conservative: 14  
Best Local Similarity: 32.32% Mismatches: 44  
Query Match: 15.13% Indels: 10  
DB: 9 Gaps: 4

US-10-020-441-2 (1-106) x US-10-140-470-221 (1-1129)

Qy 7 ArgAenGIYAepGInGInGIYlIehIshIsgInProSeRGIYValIys-----ValAla 24  
Db 267 CGCAACACCTCTGAGACGCTCCCTGTACAGAGTCTGCTGGGGGCAAGCAGCTAGTGAG 326  
Qy 25 ProGIYTYMeTProSeRGIYMeTSeRAlAArgInAArg-----ProIle 40  
Db 327 CCGGGA-----CCACAGCTATGTATGCCCCGGGTGAGGCAAGTGAAGCAACCCCTG 380  
Qy 41 AlAgiN---ThrLeuSeRGIYpHeaSpIleAlIeValIeMeTleuAlaGIImeTValAsn 59  
Db 381 TACCAAGGACGAGCCTCCAGCGCTGACGTGGCCCTGTGAGAGTGAAGCAGCAGGCC 440  
Qy 60 LeuGIInSeRGIYlIeArgValIleSeRleuProGIInProSeRAspIleProProGIY 79  
Db 441 TTCACCACTTACATCTCTCCCTCCGCTGCTGACCTGACCTCGTATCTTTGAGACGGGC 500  
Qy 80 ThrGIYValIpHeIleValIGlYTYRGIYAArgAspAspAspAspProSeRArg 98  
Db 501 ATGAACCTGCTGGTCACTGCTGGGGCAGCCCAAGTGAAGA-AGACCTCTGCCCCA 556

## RESULT 13

US-10-175-746-221  
Sequence 221, Application US/10175746  
Publication No. US2003002270A1

## GENERAL INFORMATION:

APPLICANT: Baker, Kevin P.  
APPLICANT: Beresini, Maureen  
APPLICANT: Deforge, Laura  
APPLICANT: Desnoyers, Luc  
APPLICANT: Filvaroff, Ellen  
APPLICANT: Gao, Wei-Qiang  
APPLICANT: Gerritsen, Mary E.  
APPLICANT: Goddard, Audrey  
APPLICANT: Godowski, Paul J.  
APPLICANT: Gurney, Austin L.  
APPLICANT: Sherwood, Steven  
APPLICANT: Smith, Victoria  
APPLICANT: Stewart, Timothy A.  
APPLICANT: Tumas, Daniel  
APPLICANT: Watanabe, Colin K  
APPLICANT: Wood, William  
APPLICANT: Zhang, Zemin  
TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC  
FILE REFERENCE: P330R1C353  
CURRENT APPLICATION NUMBER: US/10/175,746  
CURRENT FILING DATE: 2002-06-19  
Prior Application removed - See File Wrapper or Palm  
NUMBER OF SEQ ID NOS: 550  
SEQ ID NO 221  
LENGTH: 1129  
TYPE: DNA  
ORGANISM: Homo Sapien  
US-10-175-746-221

Alignment Scores: 0.0303 Length: 1129  
Pred. No.: 83.50 Matches: 32  
Score: 83.50 Conservative: 14  
Percent Similarity: 46.46% Mismatches: 44  
Best Local Similarity: 32.32% Indels: 10  
Query Match: 15.13% Gaps: 4

US-10-020-441-2 (1-106) x US-10-175-746-221 (1-1129)

Qy 7 ArgAenGIYAepGInGInGIYlIehIshIsgInProSeRGIYValIys-----ValAla 24

```
Db 267 CGCAACACCTCTGAGACCTCCCTGTACACAGGTCTGCTGGGGGCAAGGACGCTAGTGCGAG 326
Qy 25 ProglYTyMetProSerCysMetSerAlaArgGlnArgArg-----ProIle 40
Db 327 CCGGGA-----CCACACGCTATGTATGCCCCGGGTGAGGCGAGTGAGACCAACCCCTG 380
Qy 41 AlaGln---ThrluSerGlyPheAspIleAlaIleValMetLeuAlaGlnMetValAsn 59
Db 381 TACCAAGGGACCGGCTCCAGCGCTGACGTGGCTGTGTGAGACTGAGGCAACCAATGGCCC 440
Qy 60 LeuGlnSerGlyIleArgValIleSerLeuProGlnProSerAspIleProProProGly 79
Db 441 TTCACCAATTACATCTCCCTCCCGTGTGCTGCTGCTGACCCCTCGGTATCTTTGAGACGGGC 500
Qy 80 ThrGlyValPheIleValGlyTyTrGlyArgAspAspAspAspArgAspProSerArg 98
Db 501 ATGAACCTGCTGGGTCACTGCTGGGCGACGCCCACTGAGGA-AGACCTCTGCCCCGA 556

RESULT 14
US-10-176-918-221
; Sequence 221, Application US/10176918
; Publication No. US20030027275A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Beresini, Maureen
; APPLICANT: Deforge, Laura
; APPLICANT: Desnoyers, Luc
; APPLICANT: Filvaroff, Ellen
; APPLICANT: Gao, Wei-Qiang
; APPLICANT: Gerritsen, Mary E.
; APPLICANT: Goddard, Audrey
; APPLICANT: Gurney, Austin L.
; APPLICANT: Sherwood, Steven
; APPLICANT: Smith, Victoria
; APPLICANT: Stewart, Timothy A.
; APPLICANT: Tumas, Daniel
; APPLICANT: Watanabe, Colin K
; APPLICANT: Wood, William
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; FILE OF INVENTION: ACIDS ENCODING THE SAME
; FILE REFERENCE: P3330R1C382
; CURRENT APPLICATION NUMBER: US/10/176,918
; CURRENT FILING DATE: 2002-06-20
; Prior Application removed - See File Wrapper or Palm
; NUMBER OF SEQ ID NOS: 550
; SEQ ID NO 221
; LENGTH: 1129
; TYPE: DNA
; ORGANISM: Homo Sapien
US-10-176-918-221

Alignment Scores:
Pred. No.: 0.0303 Length: 1129
Score: 83.50 Matches: 32
Percent Similarity: 46.46% Conservative: 14
Best Local Similarity: 32.32% Mismatches: 44
Query Match: 15.13% Indels: 10
Gaps: 4

US-10-020-441-2 (1-106) x US-10-176-918-221 (1-1129)
Qy 7 ArgAsnGlyAspGlnGlnGlyIleHisGlnProSerGlyValIle-----ValAla 24
Db 267 CCACAACACCTCTGAGACGTCCTGTACACAGGTCTGCTGGGGGCAAGGACGCTAGTGCGAG 326
Qy 25 ProglYTyMetProSerCysMetSerAlaArgGlnArgArg-----ProIle 40
Db 327 CCGGGA-----CCACACGCTATGTATGCCCCGGGTGAGGCGAGTGAGACCAACCCCTG 380
Qy 41 AlaGln---ThrluSerGlyPheAspIleAlaIleValMetLeuAlaGlnMetValAsn 59
```

```
Db 381 TACCAAGGGACCGGCTCCAGCGCTGACGTGGCTGTGTGAGACTGAGGCAACCAATGGCCC 440
Qy 60 LeuGlnSerGlyIleArgValIleSerLeuProGlnProSerAspIleProProProGly 79
Db 441 TTCACCAATTACATCTCCCTCCCGTGTGCTGCTGCTGACCCCTCGGTATCTTTGAGACGGGC 500
Qy 80 ThrGlyValPheIleValGlyTyTrGlyArgAspAspAspAspArgAspProSerArg 98
Db 501 ATGAACCTGCTGGGTCACTGCTGGGCGACGCCCACTGAGGA-AGACCTCTGCCCCGA 556

RESULT 15
US-10-176-921-221
; Sequence 221, Application US/10176921
; Publication No. US20030027276A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Beresini, Maureen
; APPLICANT: Deforge, Laura
; APPLICANT: Desnoyers, Luc
; APPLICANT: Filvaroff, Ellen
; APPLICANT: Gao, Wei-Qiang
; APPLICANT: Gerritsen, Mary E.
; APPLICANT: Goddard, Audrey
; APPLICANT: Gurney, Austin L.
; APPLICANT: Sherwood, Steven
; APPLICANT: Smith, Victoria
; APPLICANT: Stewart, Timothy A.
; APPLICANT: Tumas, Daniel
; APPLICANT: Watanabe, Colin K
; APPLICANT: Wood, William
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; FILE OF INVENTION: ACIDS ENCODING THE SAME
; FILE REFERENCE: P3330R1C288
; CURRENT APPLICATION NUMBER: US/10/176,921
; CURRENT FILING DATE: 2002-06-20
; Prior Application removed - See File Wrapper or Palm
; NUMBER OF SEQ ID NOS: 550
; SEQ ID NO 221
; LENGTH: 1129
; TYPE: DNA
; ORGANISM: Homo Sapien
US-10-176-921-221

Alignment Scores:
Pred. No.: 0.0303 Length: 1129
Score: 83.50 Matches: 32
Percent Similarity: 46.46% Conservative: 14
Best Local Similarity: 32.32% Mismatches: 44
Query Match: 15.13% Indels: 10
Gaps: 4

US-10-020-441-2 (1-106) x US-10-176-921-221 (1-1129)
Qy 7 ArgAsnGlyAspGlnGlnGlyIleHisGlnProSerGlyValIle-----ValAla 24
Db 267 CCACAACACCTCTGAGACGTCCTGTACACAGGTCTGCTGGGGGCAAGGACGCTAGTGCGAG 326
Qy 25 ProglYTyMetProSerCysMetSerAlaArgGlnArgArg-----ProIle 40
Db 327 CCGGGA-----CCACACGCTATGTATGCCCCGGGTGAGGCGAGTGAGACCAACCCCTG 380
Qy 41 AlaGln---ThrluSerGlyPheAspIleAlaIleValMetLeuAlaGlnMetValAsn 59
Db 381 TACCAAGGGACCGGCTCCAGCGCTGACGTGGCTGTGTGAGACTGAGGCAACCAATGGCCC 440
Qy 60 LeuGlnSerGlyIleArgValIleSerLeuProGlnProSerAspIleProProProGly 79
Db 441 TTCACCAATTACATCTCCCTCCCGTGTGCTGCTGCTGACCCCTCGGTATCTTTGAGACGGGC 500
Qy 80 ThrGlyValPheIleValGlyTyTrGlyArgAspAspAspAspArgAspProSerArg 98
```



```

; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Sherwood, Steven
; APPLICANT: Smith, Victoria
; APPLICANT: Stewart, Timothy A.
; APPLICANT: Tumas, Daniel
; APPLICANT: Watanabe, Colin K
; APPLICANT: Wood, William
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; FILE REFERENCE: P330R1C251
; CURRENT APPLICATION NUMBER: US/10/142,431
; CURRENT FILING DATE: 2002-05-10
; Prior Application removed - See File Wrapper or Palm
; NUMBER OF SEQ ID NOS: 550
; SEQ ID NO 221
; LENGTH: 1129
; TYPE: DNA
; ORGANISM: Homo Sapien
US-10-142-431-221

Alignment Scores:
Pred. No.: 0.0303      Length: 1129
Score: 83.50      Matches: 32
Percent Similarity: 46.46%      Conservative: 14
Best Local Similarity: 32.32%      Mismatches: 44
Query Match: 15.13%      Indels: 10
DB: 9      Gaps: 4

US-10-020-441-2 (1-106) x US-10-142-431-221 (1-1129)
Qy 7 ArgAanglyAspgInglnglylleHishsgInProserglyVallys-----Valala 24
Db 267 CGCAACACCTCTGAGACGCTCCCTGTACACAGGTCCTGTGGGGGCAAGGCACTAGTGCG 326
Qy 25 ProglyTYrMetProserCyemeterAlaargInaArg-----Proile 40
Db 327 CCGGGA-----CCACAGCGTATGTATGCCCGGGTGAGGAGGAGGCAACCCCTG 380
Qy 41 AlaIn---ThrluSerglyPheaspIleAlaIleValmetLeuAlaInmetValasn 59
Db 381 TACACGGGACGCGCTCCAGCGCTGACGCGCTGCTGAGACGACGACGAGCC 440
Qy 60 LeuInSerglyIleArgValIleSerleuProglInProserAspIleProProgly 79
Db 441 TTCACCAATTAATCATCTCCCGCTGTGCTGCTGACCCCTCGGTATCTTTGAGACGGGC 500
Qy 80 ThrGlyValPheIleValGlyTYrGlyArgAspAspAspAspAspAspAspAsp 98
Db 501 ATGAACCTGCTGGGTCACTGCTGGGGGACCCCACTGAGGA-AGACCTCTGCGCGA 556

RESULT 19
US-10-143-114-221
; Sequence 221, Application US/10143114
; Publication No. US20030036180A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Beresini, Maureen
; APPLICANT: DeForge, Laura
; APPLICANT: Desnoyers, Luc
; APPLICANT: Filvaroff, Ellen
; APPLICANT: Gao, Wei-Qiang
; APPLICANT: Gerritsen, Mary E.
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Sherwood, Steven
; APPLICANT: Smith, Victoria
; APPLICANT: Stewart, Timothy A.
; APPLICANT: Tumas, Daniel
; APPLICANT: Watanabe, Colin K
; APPLICANT: Wood, William
```

```

; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; FILE REFERENCE: P330R1C211
; CURRENT APPLICATION NUMBER: US/10/143,114
; CURRENT FILING DATE: 2002-05-09
; Prior Application removed - See Palm or File Wrapper
; NUMBER OF SEQ ID NOS: 550
; SEQ ID NO 221
; LENGTH: 1129
; TYPE: DNA
; ORGANISM: Homo Sapien
US-10-143-114-221

Alignment Scores:
Pred. No.: 0.0303      Length: 1129
Score: 83.50      Matches: 32
Percent Similarity: 46.46%      Conservative: 14
Best Local Similarity: 32.32%      Mismatches: 44
Query Match: 15.13%      Indels: 10
DB: 9      Gaps: 4

US-10-020-441-2 (1-106) x US-10-143-114-221 (1-1129)
Qy 7 ArgAanglyAspgInglnglylleHishsgInProserglyVallys-----Valala 24
Db 267 CGCAACACCTCTGAGACGCTCCCTGTACACAGGTCCTGTGGGGGCAAGGCACTAGTGCG 326
Qy 25 ProglyTYrMetProserCyemeterAlaargInaArg-----Proile 40
Db 327 CCGGGA-----CCACAGCGTATGTATGCCCGGGTGAGGAGGAGGCAACCCCTG 380
Qy 41 AlaIn---ThrluSerglyPheaspIleAlaIleValmetLeuAlaInmetValasn 59
Db 381 TACACGGGACGCGCTCCAGCGCTGACGCGCTGCTGAGACGACGACGAGCC 440
Qy 60 LeuInSerglyIleArgValIleSerleuProglInProserAspIleProProgly 79
Db 441 TTCACCAATTAATCATCTCCCGCTGTGCTGCTGACCCCTCGGTATCTTTGAGACGGGC 500
Qy 80 ThrGlyValPheIleValGlyTYrGlyArgAspAspAspAspAspAspAspAsp 98
Db 501 ATGAACCTGCTGGGTCACTGCTGGGGGACCCCACTGAGGA-AGACCTCTGCGCGA 556

RESULT 20
US-10-140-002-221
; Sequence 221, Application US/10140002
; Publication No. US20030037623A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Beresini, Maureen
; APPLICANT: DeForge, Laura
; APPLICANT: Desnoyers, Luc
; APPLICANT: Filvaroff, Ellen
; APPLICANT: Gao, Wei-Qiang
; APPLICANT: Gerritsen, Mary E.
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Sherwood, Steven
; APPLICANT: Smith, Victoria
; APPLICANT: Stewart, Timothy A.
; APPLICANT: Tumas, Daniel
; APPLICANT: Watanabe, Colin K
; APPLICANT: Wood, William
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; FILE REFERENCE: P330R1C59
; CURRENT APPLICATION NUMBER: US/10/140,002
; CURRENT FILING DATE: 2002-05-06
; Prior Application removed - See Palm or File Wrapper
; NUMBER OF SEQ ID NOS: 550
```



SEQ ID NO 221  
LENGTH: 1129  
TYPE: DNA  
ORGANISM: Homo Sapien  
US-10-140-002-221

## Alignment Scores:

Pred. No.:	0.0303	Length:	1129
Score:	83.50	Matches:	32
Percent Similarity:	46.46%	Conservative:	14
Best Local Similarity:	32.32%	Mismatches:	44
Query Match:	15.13%	Indels:	10
DB:	9	Gaps:	4

US-10-020-441-2 (1-106) x US-10-140-002-221 (1-1129)

Qy 7 ArgAnGlyApBgInGlnGlyIleHISgInProSeRgLyValIlyS-----ValAla 24  
Db 267 CGCAACACCTCTGAGACGCTCCGTACACAGGTCTGCTGGGGGCAAGCAGCTAGTGCAG 326  
Qy 25 ProGlyTyMetProSeRgCyMeSeSerAlaArgGlnArg-----ProIle 40  
Db 327 CCGGGA-----CCACAGCGTATGTATGCCCCGGGTGAGGAGGTGAGAGCAACCCCTG 380  
Qy 41 AlagIn---ThrLeuSeRgLyPheApIleAlaIleValMeLeuAlaGlnMeValAsn 59  
Db 381 TACCAAGGACGGCCTCCAGCGCTGACGTGGCCTGTGAGCTGGAGGACACAGTGC 440  
Qy 60 LeuGInSeRgLyIleArgValIleSeRleuProGlnProSeRAspIleProProGly 79  
Db 441 TTCACCAATTCATCTCTCCCGGTGCTGCTGAGCCCTCGGTATCTTTGAGACGGGC 500  
Qy 80 ThrGlyValPheIleValGlyTyrcGlyArgAspAspAspAspAspProSeRArg 98  
Db 501 ATGAACCTGCTGGGTCACTGGCTGGGGCAGCCCCAGTAGGA-AGACCTCTGCCGA 556

## RESULT 21

US-10-142-419-221  
Sequence 221, Application US/10142419  
Publication No. US20030044945A1

GENERAL INFORMATION:  
APPLICANT: Baker, Kevin P.  
APPLICANT: Beresini, Maureen  
APPLICANT: DeForge, Laura  
APPLICANT: Deenoysers, Luc  
APPLICANT: Filvaroff, Ellen  
APPLICANT: Gao, Wei-Qiang  
APPLICANT: Gerritsen, Mary E.  
APPLICANT: Goddard, Audrey  
APPLICANT: Gurney, Austin L.  
APPLICANT: Godowski, Paul J.  
APPLICANT: Sherwood, Steven  
APPLICANT: Smith, Victoria  
APPLICANT: Stewart, Timothy A.  
APPLICANT: Tumas, Daniel  
APPLICANT: Watanabe, Colin K  
APPLICANT: Wood, William  
APPLICANT: Zhang, Zemin  
TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC  
FILE REFERENCE: P330R1C244  
CURRENT APPLICATION NUMBER: US/10/142,419  
CURRENT FILING DATE: 2002-05-10  
Prior Application removed - See File Wrapper or Palm  
NUMBER OF SEQ ID NOS: 550  
SEQ ID NO 221  
LENGTH: 1129  
TYPE: DNA  
ORGANISM: Homo Sapien  
US-10-142-419-221

## Alignment Scores:

Pred. No.:

0.0303

Length:

1129

Score: 83.50  
Percent Similarity: 46.46%  
Best Local Similarity: 32.32%  
Query Match: 15.13%  
DB: 9  
Matches: 32  
Conservative: 14  
Mismatches: 44  
Indels: 10  
Gaps: 4

US-10-020-441-2 (1-106) x US-10-142-419-221 (1-1129)

Qy 7 ArgAnGlyApBgInGlnGlyIleHISgInProSeRgLyValIlyS-----ValAla 24  
Db 267 CGCAACACCTCTGAGACGCTCCGTACACAGGTCTGCTGGGGGCAAGCAGCTAGTGCAG 326  
Qy 25 ProGlyTyMetProSeRgCyMeSeSerAlaArgGlnArg-----ProIle 40  
Db 327 CCGGGA-----CCACAGCGTATGTATGCCCCGGGTGAGGAGGTGAGAGCAACCCCTG 380  
Qy 41 AlagIn---ThrLeuSeRgLyPheApIleAlaIleValMeLeuAlaGlnMeValAsn 59  
Db 381 TACCAAGGACGGCCTCCAGCGCTGACGTGGCCTGTGAGCTGGAGGACACAGTGC 440  
Qy 60 LeuGInSeRgLyIleArgValIleSeRleuProGlnProSeRAspIleProProGly 79  
Db 441 TTCACCAATTCATCTCTCCCGGTGCTGCTGAGCCCTCGGTATCTTTGAGACGGGC 500  
Qy 80 ThrGlyValPheIleValGlyTyrcGlyArgAspAspAspAspAspProSeRArg 98  
Db 501 ATGAACCTGCTGGGTCACTGGCTGGGGCAGCCCCAGTAGGA-AGACCTCTGCCGA 556

## RESULT 22

US-10-123-262-221  
Sequence 221, Application US/10123262  
Publication No. US20030049816A1

GENERAL INFORMATION:  
APPLICANT: Baker, Kevin P.  
APPLICANT: Beresini, Maureen  
APPLICANT: DeForge, Laura  
APPLICANT: Deenoysers, Luc  
APPLICANT: Filvaroff, Ellen  
APPLICANT: Gao, Wei-Qiang  
APPLICANT: Gerritsen, Mary E.  
APPLICANT: Goddard, Audrey  
APPLICANT: Gurney, Austin L.  
APPLICANT: Sherwood, Steven  
APPLICANT: Smith, Victoria  
APPLICANT: Stewart, Timothy A.  
APPLICANT: Tumas, Daniel  
APPLICANT: Watanabe, Colin K  
APPLICANT: Wood, William  
APPLICANT: Zhang, Zemin  
TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC  
FILE REFERENCE: P330R1C38  
CURRENT APPLICATION NUMBER: US/10/123,262  
CURRENT FILING DATE: 2002-04-15  
Prior Application removed - See File Wrapper or Palm  
NUMBER OF SEQ ID NOS: 550  
SEQ ID NO 221  
LENGTH: 1129  
TYPE: DNA  
ORGANISM: Homo Sapien  
US-10-123-262-221

## Alignment Scores:

Pred. No.:	0.0303	Length:	1129
Score:	83.50	Matches:	32
Percent Similarity:	46.46%	Conservative:	14
Best Local Similarity:	32.32%	Mismatches:	44
Query Match:	15.13%	Indels:	10
DB:	9	Gaps:	4

US-10-020-441-2 (1-106) x US-10-123-262-221 (1-1129)





```

; CURRENT FILING DATE: 2002-01-07
; NUMBER OF SEQ ID NOS: 11
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 8
; LENGTH: 1130
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Fusion gene of
; OTHER INFORMATION: Protease T in a zymogen activation vector
US-10-041-006A-8

Alignment Scores:
Pred. No.: 0.0303 Length: 1130
Score: 83.50 Matches: 32
Percent Similarity: 46.46% Conservative: 14
Best Local Similarity: 32.32% Mismatches: 44
Query Match: 15.13% Indels: 10
DB: 9 Gaps: 4

US-10-020-441-2 (1-106) x US-10-041-006A-8 (1-1130)
QY 7 ArgAenGIyAspGInGInGIyIleHISGInProSeRGIyValIys-----ValAla 24
   |||||
   :::::
Db 295 CGCAACACCTCTGAGACGCTCCCTGTACAGGTCCTGCTGGGGGAGAGGACGACTAGTGAG 354
QY 25 ProGIyTYrMetProSeRyCysMeSeRAlAArgGInAArg-----ProIle 40
   |||||
   |||||
Db 355 CCGGGA-----CCACACGCTATGTATGCCCGGGTGAGGAGGAGGAGCAACCCCTG 408
QY 41 AlaGIn---ThrLeuSeRGIyPheAspIleAlaIleValMeLeuAlaGInMeValAen 59
   |||||
   |||||
Db 409 TACCAAGGACGCGCTCCAGCGCTGACGTCGCTGTGAGAGTGAGGACACAGGCC 468
QY 60 LeuGInSeRGIyIleAArgValIleSeRLeuProGInProSeRAspIleProProGIy 79
   :::::
   :::::
Db 469 TTCACCAATTAATCTCTCCCGCTGCTGCTGAGACCCCTCGGTGATCTTTGAGACGGGC 528
QY 80 ThrGIyValPheIleValGIyTYrGIyAArgAspAspAspAArgAspProSeRArg 98
   :::::
   :::::
Db 529 ATGAACCTGCTGGGTCACTGCTGGGGGAGCCCGAGTGAGGA-AGACCTCTGCGCGA 584

RESULT 28
US-10-040-655-8
; Sequence 8, Application US/10040655
; Patent No. US20020146805A1
; GENERAL INFORMATION:
; APPLICANT: Andrade-Gordon, Patricia
; APPLICANT: Darrow, Andrew
; APPLICANT: Qi, Jian-shen
; TITLE OF INVENTION: DNA encoding the novel human serine
; TITLE OF INVENTION: Protease T
; FILE REFERENCE: ORT-1032
; CURRENT APPLICATION NUMBER: US/10/040,655
; CURRENT FILING DATE: 2002-01-07
; NUMBER OF SEQ ID NOS: 11
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 8
; LENGTH: 1130
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Fusion gene of
; OTHER INFORMATION: Protease T in a zymogen activation vector
US-10-040-655-8

Alignment Scores:
Pred. No.: 0.0303 Length: 1130
Score: 83.50 Matches: 32
Percent Similarity: 46.46% Conservative: 14
Best Local Similarity: 32.32% Mismatches: 44
Query Match: 15.13% Indels: 10
DB: 12 Gaps: 4
```

```

US-10-020-441-2 (1-106) x US-10-040-655-8 (1-1130)
QY 7 ArgAenGIyAspGInGInGIyIleHISGInProSeRGIyValIys-----ValAla 24
   |||||
   :::::
Db 295 CGCAACACCTCTGAGACGCTCCCTGTACAGGTCCTGCTGGGGGAGAGGACGACTAGTGAG 354
QY 25 ProGIyTYrMetProSeRyCysMeSeRAlAArgGInAArg-----ProIle 40
   |||||
   |||||
Db 355 CCGGGA-----CCACACGCTATGTATGCCCGGGTGAGGAGGAGGAGCAACCCCTG 408
QY 41 AlaGIn---ThrLeuSeRGIyPheAspIleAlaIleValMeLeuAlaGInMeValAen 59
   |||||
   |||||
Db 409 TACCAAGGACGCGCTCCAGCGCTGACGTCGCTGTGAGAGTGAGGACACAGGCC 468
QY 60 LeuGInSeRGIyIleAArgValIleSeRLeuProGInProSeRAspIleProProGIy 79
   :::::
   :::::
Db 469 TTCACCAATTAATCTCTCCCGCTGCTGCTGAGACCCCTCGGTGATCTTTGAGACGGGC 528
QY 80 ThrGIyValPheIleValGIyTYrGIyAArgAspAspAspAArgAspProSeRArg 98
   :::::
   :::::
Db 529 ATGAACCTGCTGGGTCACTGCTGGGGGAGCCCGAGTGAGGA-AGACCTCTGCGCGA 584

RESULT 29
US-09-768-826-16
; Sequence 16, Application US/09768826
; Patent No. US20020012966A1
; GENERAL INFORMATION:
; APPLICANT: Shi et al.
; TITLE OF INVENTION: 18 human secreted proteins
; FILE REFERENCE: P512P1
; CURRENT APPLICATION NUMBER: US/09/768, 826
; CURRENT FILING DATE: 2001-01-25
; PRIOR APPLICATION NUMBER: PCT/US00/22350
; PRIOR FILING DATE: 2000-08-15
; PRIOR APPLICATION NUMBER: 60/148,759
; PRIOR FILING DATE: 1999-08-16
; NUMBER OF SEQ ID NOS: 61
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 16
; LENGTH: 1350
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (1135)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: SITE
; LOCATION: (1148)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: SITE
; LOCATION: (1166)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: SITE
; LOCATION: (1174)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: SITE
; LOCATION: (1181)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: SITE
; LOCATION: (1209)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: SITE
; LOCATION: (1229)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: SITE
; LOCATION: (1266)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: SITE
; LOCATION: (1285)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: SITE
; LOCATION: (1287)
```

```
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: SITE
LOCATION: (1230)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: SITE
LOCATION: (1235)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: SITE
LOCATION: (1305)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: SITE
LOCATION: (1324)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: SITE
LOCATION: (1339)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: SITE
LOCATION: (1343)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: SITE
LOCATION: (1345)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: SITE
LOCATION: (1348)
OTHER INFORMATION: n equals a,t,g, or c
US-09-768-826-16
```

```
Alignment Scores:
Pred. No.: 0.0449 Length: 1350
Score: 83.00 Matches: 33
Percent Similarity: 40.19% Conservative: 10
Best Local Similarity: 30.84% Mismatches: 28
Query Match: 15.04% Indels: 36
Gaps: 6
```

US-10-020-441-2 (1-106) x US-09-768-826-16 (1-1350)

```
QY 13 GYIIEHSHISGLNProSergly-----ValIysValAlaProGlyTyr 27
Db 652 GGGCTCCACCTTCAGACCTGCGATGAGCAGACCCCTGAGCAGCTGCGAGGCTGAGGC 711
QY 28 MetProSerGlyMetSerAlaArgGlnArg-----ProIleAlaGlnThrLeuSer 45
Db 712 GTCCCGCCGATCTGCTGCGCGCGCGCGCGCGCGAGACCCGCTGCGCGCTTCT 771
QY 46 GYIIPheApIleAlaIleValMetLeuAlaGlnMetValAsnLeuGlnSerglyIleArg 65
Db 772 GGG-----GCT 777
QY 66 ValIleSerLeuPro-----GlnProSerApIle-----Pro 76
Db 778 AGGGCGCGCTCCCGCGCGCATCAGCCTCGGAACCTTCCGCGCGAGCGACCCCGCG 837
QY 77 ProProGlyThrGlyValPheIleValGlyTyrGlyArgAspAsnAspArgAspPro 96
Db 838 CCCCCCGCATCTGCTGCGCGCTTCCGGCGGAGAGGAGGCGCGACCCCGAGGCGCGCGCG 897
QY 97 SerArgIysAsnGlyGlyIle 103
Db 898 CATCGA-----GGAGGATC 912

RESULT 30
US-10-092-004A-1
Sequence 1, Application US/10092004A
Publication No. US20030050251A1
GENERAL INFORMATION:
APPLICANT: Corvas International, Inc.
APPLICANT: Madison, Edwin L.
APPLICANT: Semple, Joseph Edward
APPLICANT: Coombs, Gary Samuel
```

```
APPLICANT: Reiner, John Eugene
APPLICANT: Ong, Edgar O.
APPLICANT: Araldi, Gian Luca
TITLE OF INVENTION: Inhibitors of Serine Protease Activity of Matricrinase or MTSPI
FILE REFERENCE: 018613/0282105
CURRENT APPLICATION NUMBER: US/10/092,004A
PRIOR FILING DATE: 2002-09-12
PRIOR APPLICATION NUMBER: PCT/US01/28137
PRIOR FILING DATE: 2001-09-07
PRIOR APPLICATION NUMBER: 09/657,986
PRIOR FILING DATE: 2000-09-08
NUMBER OF SEQ ID NOS: 11
SOFTWARE: PatentIn version 3.0
SEQ ID NO 1
LENGTH: 726
TYPE: DNA
ORGANISM: Homo Sapiens
FEATURE:
NAME/KEY: misc feature
OTHER INFORMATION: Nucleic acid sequence of protease domain of MTSPI
US-10-092-004A-1
```

```
Alignment Scores:
Pred. No.: 0.0234 Length: 726
Score: 82.50 Matches: 23
Percent Similarity: 44.71% Conservative: 15
Best Local Similarity: 27.06% Mismatches: 28
Query Match: 14.95% Indels: 19
Gaps: 3
```

US-10-020-441-2 (1-106) x US-10-092-004A-1 (1-726)

```
QY 13 GYIIEHSHISGLNProSerglyValIysValAlaProGlyTyrMetProSerGlyMet 32
Db 187 GGGCTCCAGCAGCCAGCCAG-----CGCAGCGCCCTTGGAGT----- 225
QY 33 SerAlaArgGlnArgArgProIleAlaGlnThrLeuSer----- 45
Db 226 -----CAGAGCGGAGGCTCAGCGCATCTCCACCCCTTCTCATGATGCTTC 276
QY 46 -----GlyPheApIleAlaIleValMetLeuAlaGlnMetValAsnLeuGlnSergly 63
Db 277 ACCTTCGACTATGACATCGCGCTGCGAGCTGGAGAAACCGGAGAGTACAGCTCCATG 336
QY 64 IleArgValIleSerLeuProGlnProSerAspIleProProProGlyThrGlyValPhe 83
Db 337 GTCCGCGCCCATCTGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 396
QY 84 IleValGlyTyrGly 88
Db 397 GTACGCGGCTGGGA 411

RESULT 31
US-10-099-700A-1
Sequence 1, Application US/10099700A
Publication No. US20030008372A1
GENERAL INFORMATION:
APPLICANT: Edgar O. Ong
APPLICANT: Madison, Edwin L.
TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING TRANSMEMBRANE SERINE PROTEASE 7, 1
FILE REFERENCE: 24745-1613
CURRENT APPLICATION NUMBER: US/10/099,700A
CURRENT FILING DATE: 2002-05-24
PRIOR APPLICATION NUMBER: 60/275,592
PRIOR FILING DATE: 2001-03-13
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 1
LENGTH: 3147
TYPE: DNA
ORGANISM: Homo Sapien
FEATURE:
```

```

; NAME/KEY: CDS
; LOCATION: (23)...(2589)
; OTHER INFORMATION: Nucleotide sequence encoding MTSPI
; PUBLICATION INFORMATION:
; AUTHORS: O'Brien, T.J. and Tanimoto, H.
; DATABASE ACCESSION NUMBER: Genbank #AR081724
; DATABASE ENTRY DATE: 2000-08-31
; PATENT DOCUMENT NUMBER: 5,972,616
; PATENT FILING DATE: 1998-02-20
; PUBLICATION DATE: 1999-10-26
US-10-099-700A-1

Alignment Scores:
Pred. No.: 0.16 Length: 3147
Score: 82.50 Matches: 23
Percent Similarity: 44.71% Conservative: 15
Best Local Similarity: 27.06% Mismatches: 28
Query Match: 14.95% Indels: 19
DB: 9 Gaps: 3

US-10-020-441-2 (1-106) x US-10-099-700A-1 (1-3147)
QY 13 GYIleHSHSGInProSerGlyValIysValAlaProGlyTYrMetProSerCyMet 32
DB 2051 GGCTTGACGACGACGAGCCAG-----CGCAGCGCCCTGGGGTG----- 2089
QY 33 SerAlaArgGlnArgProIleAlaGlnThrLeuSer----- 45
DB 2090 -----CAGAGCGGAGGCTCAAGCCCATCATCTCCACCCCTTTCATGACTTC 2140
QY 46 -----GlyPheAspIleAlaIleValMetLeuAlaGlnMetValAsnLeuGlnSerGly 63
DB 2141 ACCTTGACTATGACATCGCGCTGCTGGAGCTGGAGAAACGGCAGAGTACAGTCCATG 2200
QY 64 IleArgValIleSerLeuProGlnProSerAspIleProProGlyThrGlyValPhe 83
DB 2201 GTGCGGCCCATCTGCTGCCGACGCTCCCATGCTTCTTCCTGCGGCAAGGCCATCTGG 2260
QY 84 IleValGlyTYrGly 88
DB 2261 GTCACGGCTGGGGA 2275

RESULT 32
US-10-099-700A-3
; Sequence 3, Application US/10099700A
; Publication No. US20030008372A1
; GENERAL INFORMATION:
; APPLICANT: Edgar O. Ong
; APPLICANT: Edwin L. Madison
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING TRANSMEMBRANE SERINE PROTEASE 7,
; FILE REFERENCE: 24745-1613
; CURRENT APPLICATION NUMBER: US/10/099,700A
; PRIOR FILING DATE: 2002-05-24
; PRIOR APPLICATION NUMBER: 60/275,592
; NUMBER OF SEQ ID NOS: 22
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 3
; LENGTH: 3147
; TYPE: DNA
; ORGANISM: Homo Sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1865)...(2590)
; OTHER INFORMATION: Nucleic acid sequence of protease domain of MTSPI
US-10-099-700A-3

Alignment Scores:
Pred. No.: 0.16 Length: 3147
Score: 82.50 Matches: 23
Percent Similarity: 44.71% Conservative: 15
Best Local Similarity: 27.06% Mismatches: 28

```

```

Query Match: 14.95% Indels: 19
DB: 9 Gaps: 3

US-10-020-441-2 (1-106) x US-10-099-700A-3 (1-3147)
QY 13 GYIleHSHSGInProSerGlyValIysValAlaProGlyTYrMetProSerCyMet 32
DB 2051 GGCTTGACGACGACGAGCCAG-----CGCAGCGCCCTGGGGTG----- 2089
QY 33 SerAlaArgGlnArgProIleAlaGlnThrLeuSer----- 45
DB 2090 -----CAGAGCGGAGGCTCAAGCCCATCATCTCCACCCCTTTCATGACTTC 2140
QY 46 -----GlyPheAspIleAlaIleValMetLeuAlaGlnMetValAsnLeuGlnSerGly 63
DB 2141 ACCTTGACTATGACATCGCGCTGCTGGAGCTGGAGAAACGGCAGAGTACAGTCCATG 2200
QY 64 IleArgValIleSerLeuProGlnProSerAspIleProProGlyThrGlyValPhe 83
DB 2201 GTGCGGCCCATCTGCTGCCGACGCTCCCATGCTTCTTCCTGCGGCAAGGCCATCTGG 2260
QY 84 IleValGlyTYrGly 88
DB 2261 GTCACGGCTGGGGA 2275

RESULT 33
US-09-925-301-351
; Sequence 351, Application US/09925301
; Patent No. US20020052308A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
; FILE REFERENCE: PA106
; CURRENT APPLICATION NUMBER: US/09/925,301
; PRIOR FILING DATE: 2001-08-10
; PRIOR APPLICATION NUMBER: PCT/US00/05882
; PRIOR FILING DATE: 2000-03-08
; PRIOR APPLICATION NUMBER: 60/124,270
; NUMBER OF SEQ ID NOS: 12
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 351
; LENGTH: 2756
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1597)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (12540)
; OTHER INFORMATION: n equals a,t,g, or c
US-09-925-301-351

Alignment Scores:
Pred. No.: 0.254 Length: 2756
Score: 80.50 Matches: 20
Percent Similarity: 43.42% Conservative: 13
Best Local Similarity: 26.32% Mismatches: 38
Query Match: 14.58% Indels: 5
DB: 10 Gaps: 1

US-10-020-441-2 (1-106) x US-09-925-301-351 (1-2756)
QY 18 ProSerGlyValIysValAlaProGlyTYrMetProSerCyMetSerAlaArgGlnArg 37
DB 1606 CCTGGGCTTGACGACGACGAGCCAGCCCTGGGGTGCAGAGCGCAGCTCAAG 1665
QY 38 ArgProIleAlaGlnThrLeuSerGly-----PheAspIleAlaIleVal 52
DB 1666 CGCATCATCTCCACCCCTTTCATGACTTCACCTTGACTATGACATCGCGCTGCTG 1725
QY 53 MetLeuAlaGlnMetValAsnLeuGlnSerGlyIleArgValIleSerLeuProGlnPro 72

```

Db 1726 GAGCTGAGAAACCCGACAGTACAGCTCCATGCTGCGCCCATCTGCTGCGACGCC 1785  
Qy 73 SerApIleProProProGlyThrGlyValPheIleValGlyTyrGly 88  
Db 1786 TCCCATGCTTCCCTGCGCGCAAGGCATCTGGGTCAACGGGCTGGGGA 1833

## RESULT 34

US-10-041-400A-8  
; Sequence 8, Application US/10041400A  
; Patent No. US20020110895A1  
; GENERAL INFORMATION:  
; APPLICANT: Darrow, Andrew  
; APPLICANT: Andrade-Gordon, Patricia  
; APPLICANT: Qi, Jensen  
; TITLE OF INVENTION: DNA Encoding the Human Serine  
; FILE REFERENCE: Proctase EOS  
; CURRENT APPLICATION NUMBER: US/10/041,400A  
; CURRENT FILING DATE: 2002-01-08  
; PRIOR APPLICATION NUMBER: US/09/387,375  
; NUMBER OF SEQ ID NOS: 9  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 8  
; LENGTH: 1130  
; TYPE: DNA  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Description of Artificial Sequence: Nucleic acid  
US-10-041-400A-8

## Alignment Scores:

Pred. No.:	0.0927	Length:	1130
Score:	80.00	Matches:	15
Percent Similarity:	60.98%	Conservative:	10
Best Local Similarity:	36.59%	Mismatches:	16
Query Match:	14.49%	Indels:	0
DB:	12	Gaps:	0

US-10-020-441-2 (1-106) x US-10-041-400A-8 (1-1130)

Qy 48 AspIleAlaIleValMetLeuAlaGlnMetValAsnLeuGlnSerGlyIleArgValIle 67  
Db 433 GACCTGGACACTGCTCAGCTGCTGCGCGGCTGAGCCCTGAGCGTCCACACCCGTC 492  
Qy 68 SerLeuProGlnProSerAspIleProProGlyThrGlyValPheIleValGlyTyr 87  
Db 493 TGCCCTGCCGTCGCGCGCGCGCGCGCGCGCGCGCGCACACCATGCGGGTCAACCGGCTGG 552

Qy 88 Gly 88  
Db 553 GGC 555

## RESULT 35

US-10-041-264A-8  
; Sequence 8, Application US/10041264A  
; Patent No. US20020142446A1  
; GENERAL INFORMATION:  
; APPLICANT: Darrow, Andrew  
; APPLICANT: Andrade-Gordon, Patricia  
; APPLICANT: Qi, Jensen  
; TITLE OF INVENTION: DNA Encoding the Human Serine  
; FILE REFERENCE: Proctase EOS  
; CURRENT APPLICATION NUMBER: US/10/041,264A  
; CURRENT FILING DATE: 2002-01-08  
; PRIOR APPLICATION NUMBER: US/09/387,375  
; NUMBER OF SEQ ID NOS: 9  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 8

; LENGTH: 1130  
; TYPE: DNA  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Description of Artificial Sequence: Nucleic acid  
US-10-041-264A-8

## Alignment Scores:

Pred. No.:	0.0927	Length:	1130
Score:	80.00	Matches:	15
Percent Similarity:	60.98%	Conservative:	10
Best Local Similarity:	36.59%	Mismatches:	16
Query Match:	14.49%	Indels:	0
DB:	12	Gaps:	0

US-10-020-441-2 (1-106) x US-10-041-264A-8 (1-1130)

Qy 48 AspIleAlaIleValMetLeuAlaGlnMetValAsnLeuGlnSerGlyIleArgValIle 67  
Db 433 GACCTGGACACTGCTCAGCTGCTGCGCGGCTGAGCCCTGAGCGTCCACACCCGTC 492  
Qy 68 SerLeuProGlnProSerAspIleProProGlyThrGlyValPheIleValGlyTyr 87  
Db 493 TGCCCTGCCGTCGCGCGCGCGCGCGCGCGCGCGCGCACACCATGCGGGTCAACCGGCTGG 552

Qy 88 Gly 88  
Db 553 GGC 555

## RESULT 36

US-10-042-091A-8  
; Sequence 8, Application US/10042091A  
; Patent No. US20020142447A1  
; GENERAL INFORMATION:  
; APPLICANT: Darrow, Andrew  
; APPLICANT: Andrade-Gordon, Patricia  
; APPLICANT: Qi, Jensen  
; TITLE OF INVENTION: DNA Encoding the Human Serine  
; FILE REFERENCE: Proctase EOS  
; CURRENT APPLICATION NUMBER: US/10/042,091A  
; CURRENT FILING DATE: 2002-01-08  
; PRIOR APPLICATION NUMBER: US/09/387,375  
; NUMBER OF SEQ ID NOS: 9  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 8  
; LENGTH: 1130  
; TYPE: DNA  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Description of Artificial Sequence: Nucleic acid  
US-10-042-091A-8

## Alignment Scores:

Pred. No.:	0.0927	Length:	1130
Score:	80.00	Matches:	15
Percent Similarity:	60.98%	Conservative:	10
Best Local Similarity:	36.59%	Mismatches:	16
Query Match:	14.49%	Indels:	0
DB:	12	Gaps:	0

US-10-020-441-2 (1-106) x US-10-042-091A-8 (1-1130)

Qy 48 AspIleAlaIleValMetLeuAlaGlnMetValAsnLeuGlnSerGlyIleArgValIle 67  
Db 433 GACCTGGACACTGCTCAGCTGCTGCGCGGCTGAGCCCTGAGCGTCCACACCCGTC 492  
Qy 68 SerLeuProGlnProSerAspIleProProGlyThrGlyValPheIleValGlyTyr 87  
Db 493 TGCCCTGCCGTCGCGCGCGCGCGCGCGCGCGCGCGCACACCATGCGGGTCAACCGGCTGG 552

```
QY      88 GLY 88
      |||
Db      553 GGC 555

RESULT 37
US-10-041-400A-1
; Sequence 1, Application US/10041400A
; Patent No. US20020110895A1
; GENERAL INFORMATION:
; APPLICANT: Darrow, Andrew
; APPLICANT: Andrade-Gordon, Patricia
; APPLICANT: Qi, Jensen
; TITLE OF INVENTION: DNA Encoding the Human Serine
; FILE REFERENCE: ORT-1031
; CURRENT APPLICATION NUMBER: US/10/041,400A
; PRIOR FILING DATE: 2002-01-08
; PRIOR APPLICATION NUMBER: US/09/387,375
; NUMBER OF SEQ ID NOS: 9
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
; LENGTH: 1613
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-041-400A-1

Alignment Scores:
Pred. No.:      0.148      Length:      1613
Score:          80.00      Matches:      15
Percent Similarity: 60.98%      Conservative: 10
Best Local Similarity: 36.59%      Mismatches: 16
Query Match:    14.49%      Indels:      0
DB:             12          Gaps:      0

US-10-020-441-2 (1-106) x US-10-041-400A-1 (1-1613)
QY      48 AsplleAlaIleValMetLeuAlaGlnMetValAsnLeuGlnSerGlyIleArgValIle 67
      |||
Db      444 GACCTGGACACTGCTGCAGTGGCTGCCGCCGCGTGCCTGAGCGCTGCCATCAACCCGTC 503
      |||
QY      68 SerLeuProGlnProSerAspIleProProGlyThrGlyValPheIleValGlyTyr 87
      |||
Db      504 TGCCTGCCGCTGCCGCCGCCGCCGCCGCCGCCGCCGCCGCCGCCGCCGCCGCCGCCGCTGG 563
      |||
QY      88 GLY 88
      |||
Db      564 GGC 566

RESULT 38
US-10-041-264A-1
; Sequence 1, Application US/10041264A
; Patent No. US2002014246A1
; GENERAL INFORMATION:
; APPLICANT: Darrow, Andrew
; APPLICANT: Andrade-Gordon, Patricia
; APPLICANT: Qi, Jensen
; TITLE OF INVENTION: DNA Encoding the Human Serine
; FILE REFERENCE: ORT-1031
; CURRENT APPLICATION NUMBER: US/10/041,264A
; PRIOR FILING DATE: 2002-01-08
; PRIOR APPLICATION NUMBER: US/09/387,375
; NUMBER OF SEQ ID NOS: 9
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
; LENGTH: 1613
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-041-264A-1
```

```
Alignment Scores:
Pred. No.:      0.148      Length:      1613
Score:          80.00      Matches:      15
Percent Similarity: 60.98%      Conservative: 10
Best Local Similarity: 36.59%      Mismatches: 16
Query Match:    14.49%      Indels:      0
DB:             12          Gaps:      0

US-10-020-441-2 (1-106) x US-10-041-264A-1 (1-1613)
QY      48 AsplleAlaIleValMetLeuAlaGlnMetValAsnLeuGlnSerGlyIleArgValIle 67
      |||
Db      444 GACCTGGACACTGCTGCAGTGGCTGCCGCCGCGTGCCTGAGCGCTGCCATCAACCCGTC 503
      |||
QY      68 SerLeuProGlnProSerAspIleProProGlyThrGlyValPheIleValGlyTyr 87
      |||
Db      504 TGCCTGCCGCTGCCGCCGCCGCCGCCGCCGCCGCCGCCGCCGCCGCCGCCGCCGCCGCTGG 563
      |||
QY      88 GLY 88
      |||
Db      564 GGC 566

RESULT 39
US-10-042-091A-1
; Sequence 1, Application US/10042091A
; Patent No. US2002014247A1
; GENERAL INFORMATION:
; APPLICANT: Darrow, Andrew
; APPLICANT: Andrade-Gordon, Patricia
; APPLICANT: Qi, Jensen
; TITLE OF INVENTION: DNA Encoding the Human Serine
; FILE REFERENCE: ORT-1031
; CURRENT APPLICATION NUMBER: US/10/042,091A
; PRIOR FILING DATE: 2002-01-08
; PRIOR APPLICATION NUMBER: US/09/387,375
; NUMBER OF SEQ ID NOS: 9
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
; LENGTH: 1613
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-042-091A-1

Alignment Scores:
Pred. No.:      0.148      Length:      1613
Score:          80.00      Matches:      15
Percent Similarity: 60.98%      Conservative: 10
Best Local Similarity: 36.59%      Mismatches: 16
Query Match:    14.49%      Indels:      0
DB:             12          Gaps:      0

US-10-020-441-2 (1-106) x US-10-042-091A-1 (1-1613)
QY      48 AsplleAlaIleValMetLeuAlaGlnMetValAsnLeuGlnSerGlyIleArgValIle 67
      |||
Db      444 GACCTGGACACTGCTGCAGTGGCTGCCGCCGCGTGCCTGAGCGCTGCCATCAACCCGTC 503
      |||
QY      68 SerLeuProGlnProSerAspIleProProGlyThrGlyValPheIleValGlyTyr 87
      |||
Db      504 TGCCTGCCGCTGCCGCCGCCGCCGCCGCCGCCGCCGCCGCCGCCGCCGCCGCCGCCGCTGG 563
      |||
QY      88 GLY 88
      |||
Db      564 GGC 566

RESULT 40
US-09-832-197-2
; Sequence 2, Application US/09832197
; Patent No. US20020094955A1
; GENERAL INFORMATION:
; APPLICANT: Young et al.
```



/ TITLE OF INVENTION: Plasminogen-Like Polynucleotides, Polypeptides, and Antibodies  
/ FILE REFERENCE: PT019P1  
/ CURRENT APPLICATION NUMBER: US/09/832,197  
/ CURRENT FILING DATE: 2001-04-11  
/ PRIOR APPLICATION NUMBER: PCT/US00/27253  
/ PRIOR FILING DATE: 2000-10-04  
/ PRIOR APPLICATION NUMBER: 60/158,044  
/ PRIOR FILING DATE: 1999-10-07  
/ NUMBER OF SEQ ID NOS: 8  
/ SOFTWARE: PatentIn Ver. 2.0  
/ SEQ ID NO 2  
/ LENGTH: 1836  
/ TYPE: DNA  
/ ORGANISM: Homo sapiens  
US-09-832-197-2

## Alignment Scores:

Pred. No.:	0.175	Length:	1836
Score:	80.00	Matches:	15
Percent Similarity:	60.98%	Conservative:	10
Best Local Similarity:	36.59%	Mismatches:	16
Query Match:	14.49%	Indels:	0
DB:	10	Gaps:	0

US-10-020-441-2 (1-106) x US-09-832-197-2 (1-1836)

QY 48 AspIleAlaIleValMetLeuAlaGlnMetValAsnLeuGlnSerGlyIleArgValIle 67  
Db 460 GACCTGGGCACTGCTGAGCTGCGTGGCCCGGCTGAGCGCTGCGTCCACACCGCTC 519  
QY 68 SerLeuProGlnProSerAspIleProProGlyThrGlyValPheIleValGlyTyr 87  
Db 520 TGCCCTGCCCGTCCCGCGCGCCCGCGCGCCGCGCACACCATGCCCGGTCAACCGCTGG 579  
QY 88 Gly 88  
Db 580 GGC 582

Search completed: April 8, 2003, 20:51:11  
Job time : 96.623 secs

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.4 p5 4578  
Copyright (c) 1993 - 2003 Compugen Ltd.

## OM protein - nucleic search, using frame\_plus\_p2n model

Run on: April 8, 2003, 18:35:44 ; Search time 53.8689 Seconds  
(without alignments)  
603.461 Million cell updates/sec

Title: US-10-020-441-2

Perfect score: 552  
Sequence: 1 VSFLTLRNGDQGIHQPSCG.....YGRDNDNRDPFRKNGILKK 106

## Scoring table:

BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 6.0 , Ygapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 441362 seqs, 153338381 residues

Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

## Command line parameters:

-MODEL=frame+ p2n.model -DEV=x1h  
-Q=/cgn2\_1/USFTO.epool/US10020441/runat\_02042003\_093013\_21004/app\_query.fasta\_1.462  
-DB=Issued\_Patents\_NA -QPM=fastap -SUFFIX=rml -MINMATCH=0.1 -LOOPCL=0  
-LOOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=bl0sum62 -TRANS=human40.cdi  
-LIST=45 -DOCALIGN=200 -THR SCORE=DCT -THR MAX=100 -THR MIN=0 -ALIGN=40  
-MODE=LOCAL -OUTFMT=ptco -NORM=ext -HEAPSIZE=500 -MITLEN=0 -MAXLEN=2000000000  
-USER=US10020441 @CGN 1.1 47 @runat\_02042003\_093013\_21004 -NCPU=6 -ICPU=3  
-NO\_XLPRY -NO\_MMAR -LARGEQUERY -NEG SCORES=0 -WAIT -LONGLOG -DEV TIMEOUT=120  
-MARK TIMEOUT=30 -THRAUS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7  
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

## Database :

1: /cgn2\_6/ptodata/1/ina/5A.COMB.seq:\*  
2: /cgn2\_6/ptodata/1/ina/5B.COMB.seq:\*  
3: /cgn2\_6/ptodata/1/ina/6A.COMB.seq:\*  
4: /cgn2\_6/ptodata/1/ina/6B.COMB.seq:\*  
5: /cgn2\_6/ptodata/1/ina/PTUS.COMB.seq:\*  
6: /cgn2\_6/ptodata/1/ina/backfiles1.seq:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	102	18.5	1108	2	US-09-016-366A-14
2	102	18.5	1108	2	US-08-978-404B-20
3	102	18.5	2218	2	US-08-845-998-5
4	102	18.5	2218	3	US-09-206-537-5
5	102	18.5	2218	4	US-09-430-854-5
6	102	18.5	2259	2	US-08-845-998-3
7	102	18.5	2259	3	US-09-206-537-3
8	102	18.5	2259	4	US-09-430-854-3
9	102	18.5	3757	2	US-09-016-366A-13
10	102	18.5	3757	2	US-08-978-404B-19
11	101	18.3	1219	2	US-08-978-404B-7
12	100	18.1	1154	2	US-09-016-366A-16

13	100	18.1	1154	2	US-08-978-404B-11	Sequence 11, Appl
14	99	17.9	1137	2	US-09-016-366A-18	Sequence 18, Appl
15	99	17.9	1137	2	US-08-978-404B-13	Sequence 13, Appl
16	95.5	17.3	769	4	US-09-331-709-1	Sequence 1, Appl
17	93	16.8	735	4	US-09-079-970A-1	Sequence 1, Appl
18	93	16.8	771	4	US-09-079-970A-4	Sequence 4, Appl
19	93	16.8	1081	2	US-09-016-366A-22	Sequence 22, Appl
20	93	16.8	1081	2	US-08-978-404B-17	Sequence 17, Appl
21	93	16.8	1128	2	US-09-016-366A-20	Sequence 20, Appl
22	93	16.8	1128	2	US-08-978-404B-15	Sequence 15, Appl
23	91	16.5	1097	2	US-08-978-404B-4	Sequence 4, Appl
24	89	16.1	1103	2	US-09-016-366A-24	Sequence 24, Appl
25	86.5	15.7	2397	2	US-08-978-404B-2	Sequence 2, Appl
26	85.5	15.5	1095	2	US-08-978-404B-9	Sequence 9, Appl
27	83.5	15.1	654	4	US-09-280-116-57	Sequence 57, Appl
28	83.5	15.1	1109	4	US-09-088-651-6	Sequence 6, Appl
29	83.5	15.1	1110	4	US-09-386-653A-1	Sequence 1, Appl
30	83.5	15.1	1130	4	US-09-386-653A-8	Sequence 8, Appl
31	83	15.0	1031	2	US-08-978-404B-1	Sequence 1, Appl
32	82.5	14.9	3147	2	US-09-027-337-1	Sequence 1, Appl
33	82.5	14.9	3147	4	US-09-644-600-1	Sequence 1, Appl
34	82.5	14.9	3147	4	US-09-644-600-18	Sequence 18, Appl
35	81	14.7	1739	2	US-08-681-151-2	Sequence 2, Appl
36	80	14.5	2216	2	US-08-666-082B-2	Sequence 2, Appl
37	80	14.5	2219	1	US-07-882-925A-1	Sequence 1, Appl
38	80	14.5	2219	1	US-07-882-925A-2	Sequence 2, Appl
39	80	14.5	2219	1	US-08-184-012C-1	Sequence 1, Appl
40	80	14.5	2219	1	US-08-184-012C-2	Sequence 2, Appl
41	80	14.5	2232	1	US-08-334-177-1	Sequence 1, Appl
42	80	14.5	2232	5	PCT-US95-13830-1	Sequence 1, Appl
43	80	14.5	2262	1	US-07-882-925A-7	Sequence 7, Appl
44	80	14.5	2262	1	US-08-184-012C-7	Sequence 7, Appl
45	78.5	14.2	1553	4	US-09-280-116-10	Sequence 10, Appl

## ALIGNMENTS

RESULT 1  
US-09-016-366A-14  
; Sequence 14, Application US/09016366A  
; Patent No. 5953431  
; GENERAL INFORMATION:  
; APPLICANT: Stevens, Richard L.  
; APPLICANT: Huang, Chifu  
; TITLE OF INVENTION: MAST CELL PROTEASE PEPTIDE  
; TITLE OF INVENTION: INHIBITORS  
; NUMBER OF SEQUENCES: 65  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Wolf, Greenfield & Sacks, P.C.  
; STREET: 600 Atlantic Avenue  
; CITY: Boston  
; STATE: MA  
; COUNTRY: U.S.A.  
; ZIP: 02210-2211  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette  
; COMPUTER: IBM Compatible  
; OPERATING SYSTEM: DOS  
; SOFTWARE: FastSeq for Windows Version 2.0  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/016,366A  
; FILING DATE: January 30, 1998  
; CLASSIFICATION: 530  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 60/037,090  
; FILING DATE: 05-FEB-1997  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Plumer, Elizabeth R.  
; REGISTRATION NUMBER: 36,637  
; REFERENCE/DOCKET NUMBER: B0801/7093  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 617-720-3500  
; TELEFAX: 617-720-2441

```

/ TELEX:
/ INFORMATION FOR SEQ ID NO: 14:
/ SEQUENCE CHARACTERISTICS:
/   LENGTH: 1108 base pairs
/   TYPE: nucleic acid
/   STRANDEDNESS: single
/   TOPOLOGY: linear
/   MOLECULE TYPE: cDNA
/   US-09-016-366A-14

Alignment Scores:
Pred. No.: 0.00234      Length: 1108
Score: 102.00          Matches: 21
Percent Similarity: 62.50%      Conservative: 9
Best Local Similarity: 43.75%    Mismatches: 18
Query Match: 18.48%           Indels: 0
DB: 2                    Gaps: 0

US-10-020-441-2 (1-106) x US-09-016-366A-14 (1-1108)

Qy 46 G1YPheAp1leAla1leValmetLeuAlaGlnmetValAsnLeuGlnSerGly1leArg 65
    |||.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|
Db 391 GGGGAGACGTTGCCCTGTGAGCTTGAGTCCTGTGAATGTCCTCCACCATTCCAC 450

Qy 66 Val1leSerLeuProGlnProSerAsp1leProProGly1ThrGlyValPhe1leVal 85
    |||.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|
Db 451 CCCATATCCCTGCCCTGCGAGACCTTCCCCCTGGAGACATCGTGTGGTGACA 510

Qy 86 G1YTYrG1YrGAspAspAsnAsp 93
    |||.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|
Db 511 GGCTGGGGGACATTGATTAATGAC 534

RESULT 2
US-08-978-404B-20
/ Sequence 20, Application US/08978404B
/ Patent No. 5968782
/ GENERAL INFORMATION:
/   APPLICANT: Stevens, Richard L.
/   TITLE OF INVENTION: MAST CELL PROTEASE THAT CLEAVES
/   TITLE OF INVENTION: FIBRINOGEN
/   NUMBER OF SEQUENCES: 74
/   CORRESPONDENCE ADDRESS:
/     ADDRESS: Wolf, Greenfield & Sacks, P.C.
/     STREET: 600 Atlantic Avenue
/     CITY: Boston
/     STATE: MA
/     COUNTRY: U.S.A.
/     ZIP: 02210-2211
/   COMPUTER READABLE FORM:
/     MEDIUM TYPE: Diskette
/     COMPUTER: IBM Compatible
/     OPERATING SYSTEM: DOS
/   SOFTWARE: FASTSEQ for Windows Version 2.0
/   CURRENT APPLICATION DATA:
/     APPLICATION NUMBER: US/08/978.404B
/     FILING DATE: 25-NOV-97
/     CLASSIFICATION: 435
/     PRIOR APPLICATION DATA:
/       APPLICATION NUMBER: 60/032.354
/       FILING DATE: 04-DEC-1996
/       ATTORNEY/AGENT INFORMATION:
/         NAME: Plumer, Elizabeth R.
/         REGISTRATION NUMBER: 36.637
/         REFERENCE/DOCKET NUMBER: B0801/7090
/         TELECOMMUNICATION INFORMATION:
/           TELEPHONE: 617-720-3500
/           TELEFAX: 617-720-2441
/       TELEX:
/         INFORMATION FOR SEQ ID NO: 20:
/         SEQUENCE CHARACTERISTICS:
/           LENGTH: 1108 base pairs
/           TYPE: nucleic acid
/           STRANDEDNESS: single

```

```

/ TOPOLOGY: linear
/ US-08-978-404B-20

Alignment Scores:
Pred. No.: 0.00234      Length: 1108
Score: 102.00          Matches: 21
Percent Similarity: 62.50%      Conservative: 9
Best Local Similarity: 43.75%    Mismatches: 18
Query Match: 18.48%           Indels: 0
DB: 2                    Gaps: 0

US-10-020-441-2 (1-106) x US-08-978-404B-20 (1-1108)

Qy 46 G1YPheAp1leAla1leValmetLeuAlaGlnmetValAsnLeuGlnSerGly1leArg 65
    |||.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|
Db 391 GGGGAGACGTTGCCCTGTGAGCTTGAGTCCTGTGAATGTCCTCCACCATTCCAC 450

Qy 66 Val1leSerLeuProGlnProSerAsp1leProProGly1ThrGlyValPhe1leVal 85
    |||.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|
Db 451 CCCATATCCCTGCCCTGCGAGACCTTCCCCCTGGAGACATCGTGTGGTGACA 510

Qy 86 G1YTYrG1YrGAspAspAsnAsp 93
    |||.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|
Db 511 GGCTGGGGGACATTGATTAATGAC 534

RESULT 3
US-08-845-998-5
/ Sequence 5, Application US/08845998
/ Patent No. 5879892
/ GENERAL INFORMATION:
/   APPLICANT: Van Baren, Nicolas
/   APPLICANT: Coulie, Pierre G.
/   APPLICANT: De Smet, Charles
/   APPLICANT: Lucas, Sophie
/   APPLICANT: Boon, Thierry
/   TITLE OF INVENTION: LEUKEMIA ASSOCIATED GENES
/   NUMBER OF SEQUENCES: 16
/   CORRESPONDENCE ADDRESS:
/     ADDRESS: Wolf, Greenfield & Sacks, P.C.
/     STREET: 600 Atlantic Avenue
/     CITY: Boston
/     STATE: MA
/     COUNTRY: US
/     ZIP: 02210
/   COMPUTER READABLE FORM:
/     MEDIUM TYPE: Floppy disk
/     COMPUTER: IBM PC compatible
/     OPERATING SYSTEM: PC-DOS/MS-DOS
/     SOFTWARE: Patent Release #1.0, Version #1.25
/   CURRENT APPLICATION DATA:
/     APPLICATION NUMBER: US/08/845.998
/     FILING DATE:
/     CLASSIFICATION: 435
/     ATTORNEY/AGENT INFORMATION:
/       NAME: Van Amsterdam, John R.
/       REGISTRATION NUMBER: 40.212
/       REFERENCE/DOCKET NUMBER: I0461/7008
/       TELECOMMUNICATION INFORMATION:
/         TELEPHONE: (617)720-3500
/         TELEFAX: (617)720-2441
/       INFORMATION FOR SEQ ID NO: 5:
/       SEQUENCE CHARACTERISTICS:
/         LENGTH: 2218 base pairs
/         TYPE: nucleic acid
/         STRANDEDNESS: double
/         TOPOLOGY: linear
/         MOLECULE TYPE: cDNA
/         HYPOTHETICAL: NO
/         ANTI-SENSE: NO
/         IMMEDIATE SOURCE:
/         CLONE: NB352/3
/         FEATURE:
/           NAME/KEY: CDS

```



```
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/430,854
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/845,998
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Van Amsterdam, John R.
REGISTRATION NUMBER: 40,212
REFERENCE/DOCKET NUMBER: L0461/7008
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617)720-3500
TELEFAX: (617)720-2441
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 2218 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: cDNA
HYPOTHETICAL: NO
ANTI-SENSE: NO
IMMEDIATE SOURCE:
CLONE: N9352/3
FEATURE:
NAME/KEY: CDS
LOCATION: 8..577
US-09-430-854-5
```

```
Alignment Scores:
Pred. No.: 0.00583 Length: 2218
Score: 102.00 Matches: 35
Percent Similarity: 48.57% Conservative: 16
Best Local Similarity: 33.33% Mismatches: 35
Query Match: 18.48% Indels: 19
DB: Gaps: 4
```

US-10-020-441-2 (1-106) x US-09-430-854-5 (1-2218)

```
Qy 9 G1yApgInGInG1y1LeH1eH1eGInPProSerG1y-----Vallys 22
Db 223 GCGGACATCGCTGGAGACCGGACATCAAGATCTGCCGCCCTCAGGGTGCACCTCGGGA 282
Qy 23 Val1aPProG1yTyMePProSerCyMeSeSer1aArGInArGArGPro1le1aGIn 42
Db 283 GCAGACCT---CTACTACAGACCAAGCTGCTGCGGTCAAGACGATCATGTCACCC 339
Qy 43 Th1eu-----SerG1yPheAsP1le1a1leValMe1eua1aGInMe1Va 58
Db 340 ACAgTTCTACATCATCCAGACCGGGCGGACATCGCCCTGCTGGAGCTGGAGACCCGCT 399
Qy 58 1aSn1euGInSerG1y1leArGVal1leSe1euPProGInPProSerAsP1leProProPr 78
Db 400 GAACATCTCCAGCCACATCCACAGCGTCAAGCTGCCCTCGCTCGAGACCTTCCCCC 459
Qy 78 cG1yTh1rG1yValPhe1leValG1yTyG1yArGAsPAsPAsPAsPAsPAsPAsPAsP 98
Db 460 GGGGATGCCGTCGCTGGCTGCTGCTGGGGCGACGCTGACAT----- 502
Qy 98 g1yAsnG1yG1y 102
Db 503 ---AATGTTGG 511
```

RESULT 6  
US-08-845-998-3  
Sequence 3, Application US/08845998  
Patent No. 5879892  
GENERAL INFORMATION:  
APPLICANT: Van Baren, Nicolaas

```
APPLICANT: Coulie, Pierre G.
APPLICANT: De Smet, Charles
APPLICANT: Lucas, Sophie
APPLICANT: Boon, Thierry
TITLE OF INVENTION: LEUKEMIA ASSOCIATED GENES
NUMBER OF SEQUENCES: 16
CORRESPONDENCE ADDRESS:
ADDRESSEE: Wolf, Greenfield & Sacks, P.C.
STREET: 600 Atlantic Avenue
CITY: Boston
STATE: MA
COUNTRY: US
ZIP: 02210
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/845,998
FILING DATE:
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Van Amsterdam, John R.
REGISTRATION NUMBER: 40,212
REFERENCE/DOCKET NUMBER: L0461/7008
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617)720-3500
TELEFAX: (617)720-2441
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 2259 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: cDNA
HYPOTHETICAL: NO
ANTI-SENSE: NO
FEATURE:
NAME/KEY: CDS
LOCATION: 8..577
US-08-845-998-3
```

```
Alignment Scores:
Pred. No.: 0.00597 Length: 2259
Score: 102.00 Matches: 35
Percent Similarity: 48.57% Conservative: 16
Best Local Similarity: 33.33% Mismatches: 35
Query Match: 18.48% Indels: 19
DB: Gaps: 4
```

US-10-020-441-2 (1-106) x US-08-845-998-3 (1-2259)

```
Qy 9 G1yApgInGInG1y1LeH1eH1eGInPProSerG1y-----Vallys 22
Db 223 GCGGACATCGCTGGAGACCGGACATCAAGATCTGCCGCCCTCAGGGTGCACCTCGGGA 282
Qy 23 Val1aPProG1yTyMePProSerCyMeSeSer1aArGInArGArGPro1le1aGIn 42
Db 283 GCAGACCT---CTACTACAGACCAAGCTGCTGCGGTCAAGACGATCATGTCACCC 339
Qy 43 Th1eu-----SerG1yPheAsP1le1a1leValMe1eua1aGInMe1Va 58
Db 340 ACAgTTCTACATCATCCAGACCGGGCGGACATCGCCCTGCTGGAGCTGGAGACCCGCT 399
Qy 58 1aSn1euGInSerG1y1leArGVal1leSe1euPProGInPProSerAsP1leProProPr 78
Db 400 GAACATCTCCAGCCACATCCACAGCGTCAAGCTGCCCTCGCTCGAGACCTTCCCCC 459
Qy 78 cG1yTh1rG1yValPhe1leValG1yTyG1yArGAsPAsPAsPAsPAsPAsPAsPAsP 98
Db 460 GGGGATGCCGTCGCTGGCTGCTGCTGGGGCGACGCTGACAT----- 502
```

QY 98 glysaasnglygly 102  
Db 503 ----AATGTTGGG 511

RESULT 7  
US-09-206-537-3  
Sequence 3, Application US/09206537  
Patent No. 6130052  
GENERAL INFORMATION:  
APPLICANT: Van Baren, Nicolas  
APPLICANT: Coulie, Pierre G.  
APPLICANT: De Smet, Charles  
APPLICANT: Lucas, Sophie  
APPLICANT: Boon, Thierry  
TITLE OF INVENTION: LEUKEMIA ASSOCIATED GENES  
NUMBER OF SEQUENCES: 16  
CORRESPONDENCE ADDRESSES:  
ADDRESSEE: Wolf, Greenfield & Sacks, P.C.  
STREET: 600 Atlantic Avenue  
CITY: Boston  
STATE: MA  
COUNTRY: US  
ZIP: 02210  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/206,537  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/845,998  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: Van Amsterdam, John R.  
REGISTRATION NUMBER: 40,212  
REFERENCE/DOCKET NUMBER: L0461/7008  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (617)720-3500  
TELEFAX: (617)720-2441  
INFORMATION FOR SEQ ID NO: 3:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 2259 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
MOLECULE TYPE: CDNA  
HYPOTHETICAL: NO  
ANTI-SENSE: NO  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 8..577  
US-09-206-537-3

Alignment Scores:  
Pred. No.: 0.00597 Length: 2259  
Score: 102.00 Matches: 35  
Percent Similarity: 48.57% Conservative: 16  
Best Local Similarity: 33.33% Mismatches: 35  
Query Match: 18.48% Indels: 19  
DB: 3 Gaps: 4

US-10-020-441-2 (1-106) x US-09-206-537-3 (1-2259)

QY 9 glysaasnglygly 102  
Db 223 GGGCAGCTGCGTGAACCGGATCTGCGCCCTCCAGGAGTCACTGCGGGA 282

QY 23 Valiaaproglytyrmetprosercysmetseralarginalargyproilaagln 42  
Db 283 GCACGACCT---CTACTACGACGACCAAGCTGCTGCGGTGACAGAGATCATGTCACACC 339

QY 43 Thrleu-----Serglypheaspillealailevalmetleualaglnmetva 58  
Db 340 ACAGTTCTACATCATCCAGACGGGGCGGACATCGCCCTGCTGAGAGTGGAGAGCCCGT 399

QY 58 lasnleuglnserglylleargvalileserleuproglnproseraspillepropropr 78  
Db 400 GAACATCTCCAGCCACATCCACAGGTCAAGCTGCGCCCTGCTGCGAGACCTTCCCGCC 459

QY 78 cgllythrcllyvalphelevalglytyrglyargaspasnaspsarargaproserar 98  
Db 460 GGGATGCGGTGCTGGTCACTGCTGGGGGACGTGACAT----- 502

QY 98 glysaasnglygly 102  
Db 503 ----AATGTTGGG 511

RESULT 8  
US-09-430-854-3  
Sequence 3, Application US/09430854  
Patent No. 6271019  
GENERAL INFORMATION:  
APPLICANT: Van Baren, Nicolas  
APPLICANT: Coulie, Pierre G.  
APPLICANT: De Smet, Charles  
APPLICANT: Lucas, Sophie  
APPLICANT: Boon, Thierry  
TITLE OF INVENTION: LEUKEMIA ASSOCIATED GENES  
NUMBER OF SEQUENCES: 16  
CORRESPONDENCE ADDRESSES:  
ADDRESSEE: Wolf, Greenfield & Sacks, P.C.  
STREET: 600 Atlantic Avenue  
CITY: Boston  
STATE: MA  
COUNTRY: US  
ZIP: 02210  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/430,854  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/845,998  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: Van Amsterdam, John R.  
REGISTRATION NUMBER: 40,212  
REFERENCE/DOCKET NUMBER: L0461/7008  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (617)720-3500  
TELEFAX: (617)720-2441  
INFORMATION FOR SEQ ID NO: 3:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 2259 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
MOLECULE TYPE: CDNA  
HYPOTHETICAL: NO  
ANTI-SENSE: NO  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 8..577  
US-09-430-854-3

Alignment Scores:  
Pred. No.: 0.00597 Length: 2259  
Score: 102.00 Matches: 35  
Percent Similarity: 48.57% Conservative: 16

Best Local Similarity: 33.33% Mismatches: 35  
 Query Match: 18.48% Indels: 19  
 DB: 4 Gaps: 4

US-10-020-441-2 (1-106) x US-09-430-854-3 (1-2259)

QY 9 G1YAspGInGly1Le1eH1eH1eG1nProSerGly1-----Val1ys 22  
 |||:::|||||  
 Db 223 GGGCGACGCGTGGAAACCGGACATCAAGATCTGGCCGCCCTCAAGGTGCAACTCGGGA 282

QY 23 Val1AlaProGly1YrMeProSerCyseMetSer1AlaG1nArgArgPro1le1a1n 42  
 |||:::|||||  
 Db 283 GCAGACCT---CTACTCCAGGACCAAGCTGCTGCCGCTCGAGATCATGTCACCCC 339

QY 43 Thr1eu-----SerGlyPheAsp1leAla1leValMe1eua1aG1nMetVa 58  
 |||:::|||||  
 Db 340 ACAGTTCATCATCATCCAGACCGGGGCGGACATCCCTCTGCTGGAGCTGGAGGACCCGT 399

QY 58 1Asn1euG1nSerGly1LeArgVal1leSer1euProG1nProSerAsp1leProPr 78  
 |||:::|||||  
 Db 400 GAACATCTCCAGCCACATCCACAGCGTCAAGCTGCCCTGCTCGAGACCTTCCCCC 459

QY 78 CG1YThrGlyValPhe1leVal1Gly1YrGlyArgAspAspAsnAspArgAspProSerAr 98  
 |||:::|||||  
 Db 460 GGGGATGCCGTGCTGGTCACTGCTGGGGCGACGTGACAT----- 502

QY 98 GlyAsnGlyGly 102  
 |||:::|||||  
 Db 503 ----ATGTTGG 511

RESULT 9  
 US-09-016-366A-13  
 ; Sequence 13, Application US/09016366A  
 ; Patent No. 5955431

GENERAL INFORMATION:  
 APPLICANT: Stevens, Richard L.  
 APPLICANT: Huang, Chifu  
 TITLE OF INVENTION: MAST CELL PROTEASE PEPTIDE  
 TITLE OF INVENTION: INHIBITORS  
 NUMBER OF SEQUENCES: 65

CORRESPONDENCE ADDRESS:  
 ADDRESSEE: Wolf, Greenfield & Sacks, P.C.  
 STREET: 600 Atlantic Avenue  
 CITY: Boston  
 STATE: MA  
 COUNTRY: U.S.A.  
 ZIP: 02210-2211

COMPUTER READABLE FORM:  
 MEDIUM TYPE: Diskette  
 COMPUTER: IBM Compatible  
 OPERATING SYSTEM: DOS  
 SOFTWARE: FastSeq for Windows Version 2.0

CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/09/016,366A  
 FILING DATE: January 30, 1998  
 CLASSIFICATION: 530  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: 60/037,090  
 FILING DATE: 05-FEB-1997  
 ATTORNEY/AGENT INFORMATION:  
 NAME: Plumer, Elizabeth R.  
 REGISTRATION NUMBER: 36,637  
 REFERENCE/DOCKET NUMBER: B0801/7093  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: 617-720-3500  
 TELEFAX: 617-720-2441  
 TELEX:

INFORMATION FOR SEQ ID NO: 13:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 3757 base pairs  
 TYPE: nucleic acid  
 STRANDEDNESS: single  
 TOPOLOGY: linear

MOLECULE TYPE: cDNA  
 US-09-016-366A-13

Alignment Scores:  
 Pred. No.: 0.0117 Length: 3757  
 Score: 102.00 Matches: 21  
 Percent Similarity: 62.50% Conservative: 9  
 Best Local Similarity: 43.75% Mismatches: 18  
 Query Match: 18.48% Indels: 0  
 DB: 2 Gaps: 0

US-10-020-441-2 (1-106) x US-09-016-366A-13 (1-3757)

QY 46 G1YpheaSp1leAla1leValMe1eua1aG1nMetVal1Asn1euG1nSerGly1LeArg 65  
 |||:::|||||  
 Db 2419 GGGGAGACGTTGGCCCTGCTGGAGCTTGAGTCCCTGTGATATGTCACCAATATCCAC 2478

QY 66 Val1leSer1euProG1nProSerAsp1leProProGly1ThrGlyValPhe1leVal 85  
 |||:::|||||  
 Db 2479 CCATATCTCTGCCCTGCTCGAGACCTTCCCTCGGAGATCATGCTGGGTGACA 2538

QY 86 G1YTyrglyArgAspAspAsp 93  
 |||:::|||||  
 Db 2539 GGGTGGGGCGACATTGATGATGAC 2562

RESULT 10  
 US-08-978-404B-19  
 ; Sequence 19, Application US/08978404B  
 ; Patent No. 5968782

GENERAL INFORMATION:  
 APPLICANT: Stevens, Richard L.  
 TITLE OF INVENTION: MAST CELL PROTEASE THAT CLEAVES  
 TITLE OF INVENTION: FIBRINOGEN  
 NUMBER OF SEQUENCES: 74

CORRESPONDENCE ADDRESS:  
 ADDRESSEE: Wolf, Greenfield & Sacks, P.C.  
 STREET: 600 Atlantic Avenue  
 CITY: Boston  
 STATE: MA  
 COUNTRY: U.S.A.  
 ZIP: 02210-2211

COMPUTER READABLE FORM:  
 MEDIUM TYPE: Diskette  
 COMPUTER: IBM Compatible  
 OPERATING SYSTEM: DOS  
 SOFTWARE: FastSeq for Windows Version 2.0

CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/08/978,404B  
 FILING DATE: 25-NOV-97  
 CLASSIFICATION: 435  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: 60/032,354  
 FILING DATE: 04-DEC-1996  
 ATTORNEY/AGENT INFORMATION:  
 NAME: Plumer, Elizabeth R.  
 REGISTRATION NUMBER: 36,637  
 REFERENCE/DOCKET NUMBER: B0801/7090  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: 617-720-3500  
 TELEFAX: 617-720-2441  
 TELEX:

INFORMATION FOR SEQ ID NO: 19:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 3757 base pairs  
 TYPE: nucleic acid  
 STRANDEDNESS: single  
 TOPOLOGY: linear

US-08-978-404B-19

Alignment Scores:  
 Pred. No.: 0.0117 Length: 3757  
 Score: 102.00 Matches: 21  
 Percent Similarity: 62.50% Conservative: 9



```

US-10-020-441-2 (1-106) x US-08-978-404B-19 (1-3757)
Qy 46 Glypheaspllelaiailevalmetleualaglnmetvalaslnleuglnserglyllearg 65
Db 2419 GGGGCGAGCGTGGCCCTGCTGAGACTGAGACGTCCCTGTAAATGTCACCCCATATCCAC 2478
Qy 66 Valileserleuproglnproseraspilleproproglythrcglyvalpheleleval 85
Db 2479 CCCATATCCCTGGCCCCCTGCTGAGAGACCTTCCCCCTGGAGCATCGTCGGGTGACA 2538
Qy 86 Glytyrqllyargapapapanaap 93
Db 2539 GGCTGGGGCGACATTGATATATGAC 2562

RESULT 11
US-08-978-404B-7
; Sequence 7, Application US/08978404B
; Patent No. 5968782
; GENERAL INFORMATION:
; APPLICANT: Stevens, Richard L.
; TITLE OF INVENTION: MAST CELL PROTEASE THAT CLEAVES
; TITLE OF INVENTION: FIBRINOGEN
; NUMBER OF SEQUENCES: 74
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Wolf, Greenfield & Sacke, P.C.
; STREET: 600 Atlantic Avenue
; CITY: Boston
; STATE: MA
; COUNTRY: U.S.A.
; ZIP: 02210-2211
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/978,404B
; FILING DATE: 25-NOV-97
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 60/032,354
; FILING DATE: 04-DEC-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Plumer, Elizabeth R.
; REGISTRATION NUMBER: 36,637
; REFERENCE/DOCKET NUMBER: B0801/7090
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617-720-3500
; TELEFAX: 617-720-2441
; TELEX:
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1219 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
US-08-978-404B-7

Alignment Scores:
Pred. NO.: 0.0035 Length: 1219
Score: 101.00 Matches: 21
Percent Similarity: 60.78% Conservative: 10
Best Local Similarity: 41.18% Mismatches: 20
Query Match: 18.30% Indels: 0
DB: Gaps: 0

US-10-020-441-2 (1-106) x US-08-978-404B-7 (1-1219)
Qy 43 Thrieuserglypheaspllelaiailevalmetleualaglnmetvalaslnleuglnser 62

```

[illegible]

```

Db 314 CCAAGTGTGTCGACGACGACAGATCATGTGCACCAACGTTTCAATCATTCAGACTGG 373
Qy 46 yPheapSllleaIlaIleValMeLleuAlaGlnMetValAsnLeuGlnSergIyleatgVa 66
Db 374 AGCGGATATCCGCCCTGTGCGAGAGACCCCGTGAACTATCCAGCGCGCTCCACAC 433
Qy 66 lIleSerleuProGlnInProSerapApIleProProGlyThyrGlyValPheIleValGl 86
Db 434 GGTCAATGTGCGCCCTGTGCGAGACCTTCCGCCGGGAGTGCCTGTGGGTCACTGG 493
Qy 86 yTyrGlyArGAspAspAsnAp 93
Db 494 CTGGGGCGATGTGGACATGAT 515

RESULT 13
US-08-978-404B-11
; Sequence 11, Application US/08978404B
; Patent No. 5968782
; GENERAL INFORMATION:
; APPLICANT: Stevens, Richard L.
; TITLE OF INVENTION: MAST CELL PROTEASE THAT CLEAVES
; TITLE OF INVENTION: FLBRINOGEN
; NUMBER OF SEQUENCES: 74
; CORRESPONDENCE ADDRESSES:
; ADDRESSEE: Wolf, Greenfield & Sacks, P.C.
; STREET: 600 Atlantic Avenue
; CITY: Boston
; STATE: MA
; COUNTRY: U.S.A.
; ZIP: 02210-2211
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/978,404B
; FILING DATE: 25-NOV-97
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 60/032,354
; FILING DATE: 04-DEC-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Plummer, Elizabeth R.
; REGISTRATION NUMBER: 36,637
; REFERENCE/DOCKET NUMBER: B0801/7090
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617-720-3500
; TELEFAX: 617-720-2441
; TELEX:
; INFORMATION FOR SEQ ID NO: 11:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1154 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; US-08-978-404B-11

Alignment Scores:
Pred. No.: 0.0043 Length: 1154
Score: 100.00 Matches: 28
Percent Similarity: 50.00% Conservative: 16
Best Local Similarity: 31.82% Mismatches: 39
Query Match: 18.12% Indels: 5
DB: 2 Gaps: 1

US-10-020-441-2 (1-1106) x US-08-978-404B-11 (1-1154)
Qy 11 GlnGlnGlyIleHisIleGlnInProSergIyValIysValAlaProGlyTyrMetProSer 30
Db 254 CGTCAAAGATCTGGGCACCCCTGAGGTGCACCTGCGGGAGGACGACACCTCTATACAGGA 311
Qy 31 CysMetSerAlaIyGlnIyArGlyrProIleAlaGlnIhrIeu-----SergI 46

```

[illegible]

Db 351 GGAGCGCATCGCCCTGAGCTGAGGAGCGGATGAACTCTCCAGCCACGCTCCAC 410  
Qy 66 ValleserleuProGlnProSerapilleProProGlyThyGlyValPheIleVal 85  
Db 411 ACGGTACCTGCGCCCTGCTGAGACCTTCCCGGGAGTCCGCTGCTGCTACT 470  
Qy 86 G1YTYRG1YArgApAspAspAsp 93  
Db 471 GGCTGGGCGATGTGACATGAT 494

RESULT 15  
US-08-978-404B-13  
; Sequence 13, Application US/08978404B  
; Patent No. 5968782  
; GENERAL INFORMATION:  
; APPLICANT: Stevens, Richard L.  
; TITLE OF INVENTION: MAST CELL PROTEASE THAT CLEAVES  
; NUMBER OF INVENTION: FIBRINOGEN  
; NUMBER OF SEQUENCES: 74  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Wolf, Greenfield & Sacks, P.C.  
; STREET: 600 Atlantic Avenue  
; CITY: Boston  
; STATE: MA  
; COUNTRY: U.S.A.  
; ZIP: 02210-2211  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette  
; COMPUTER: IBM Compatible  
; OPERATING SYSTEM: DOS  
; SOFTWARE: FastSeq for Windows Version 2.0  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/978,404B  
; FILING DATE: 25-NOV-97  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 60/032,354  
; FILING DATE: 04-DEC-1996  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Plumer, Elizabeth R.  
; REGISTRATION NUMBER: 36,637  
; REFERENCE/DOCKET NUMBER: B0801/7090  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 617-720-3500  
; TELEFAX: 617-720-2441  
; TELEX:  
; INFORMATION FOR SEQ ID NO: 13:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1137 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; US-08-978-404B-13

Alignment Scores:  
Pred. No.: 0.00557 Length: 1137  
Score: 99.00 Matches: 19  
Percent Similarity: 62.50% Conservative: 11  
Best Local Similarity: 39.58% Mismatches: 18  
Query Match: 17.93% Indels: 0  
DB: 2 Gaps: 0

US-10-020-441-2 (1-106) x US-08-978-404B-13 (1-1137)

Qy 46 GlyPheAspIleAlaIleValMetLeuAlaGlnMetValAsnLeuGlnSerGlyIleArg 65  
Db 351 GGAGCGCATCGCCCTGAGCTGAGGAGCGGATGAACTCTCCAGCCACGCTCCAC 410  
Qy 66 ValleserleuProGlnProSerapilleProProGlyThyGlyValPheIleVal 85  
Db 411 ACGGTACCTGCGCCCTGCTGAGACCTTCCCGGGAGTCCGCTGCTGCTACT 470  
Qy 86 G1YTYRG1YArgApAspAspAsp 93

Db 471 GGCTGGGCGATGTGACATGAT 494

RESULT 16  
US-09-331-709-1  
; Sequence 1, Application US/09331709  
; Patent No. 6248530  
; GENERAL INFORMATION:  
; APPLICANT: Naari, Nahas  
; APPLICANT: Dumas, Milne, Edward, Jean-Baptiste  
; TITLE OF INVENTION: METHOD FOR ELIMINATING SPECIFIC  
; TITLE OF INVENTION: SEQUENCES WHEN CONSTRUCTING DNA LIBRARIES  
; FILE REFERENCE: GENSET.065APC  
; CURRENT APPLICATION NUMBER: US/09/331,709  
; PRIOR FILING DATE: 1999-07-23  
; PRIOR APPLICATION NUMBER: France 96 15854  
; NUMBER OF SEQ ID NOS: 4  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 1  
; LENGTH: 769  
; TYPE: DNA  
; ORGANISM: Homo sapien  
; FEATURE:  
; NAME/KEY: misc.feature  
; LOCATION: (1)...(769)  
; OTHER INFORMATION: n = A,T,C or G  
US-09-331-709-1

Alignment Scores:  
Pred. No.: 0.00882 Length: 769  
Score: 95.50 Matches: 25  
Percent Similarity: 47.67% Conservative: 16  
Best Local Similarity: 29.07% Mismatches: 34  
Query Match: 17.30% Indels: 11  
DB: 4 Gaps: 1

US-10-020-441-2 (1-106) x US-09-331-709-1 (1-769)

Qy 4 LeuThrLeuArgAsnGlyAspGlnGlnGlyIleHisGlnProSerGlyValIleVal 23  
Db 281 CTGCTGTGAAGAGGCGCCGAGAGGTATCCCATCACTGAGAGCTGTTGTG 340  
Qy 24 AlaProGlyIleMetProSerCysMetSerAlaArgGlnArgArgProIleAlaGlnThr 43  
Db 341 CATCCACTGTGAACCGCTGCTGTGCTGT----- 373  
Qy 44 LeuSerGlyPheAspIleAlaIleValMetLeuAlaGlnMetValAsnLeuGlnSerGly 63  
Db 374 -----GGCATGACATGCGCCCTCATCAAGCTTCACGACAGCGCCAGCTGGAGATGCC 427  
Qy 64 IleArgValIleSerleuProGlnProSerapilleProProGlyThyGlyValPhe 83  
Db 428 GTCCAGCTGCGCTCATCTCCCGCTGTGATCATCTTCCCAACAGACACCTGTGTAC 487  
Qy 84 IleValGlyTYRG1YArg 89  
Db 488 ATCACCGGCTGGGGCGGT 505

RESULT 17  
US-09-079-970A-1  
; Sequence 1, Application US/09079970A  
; Patent No. 6274366  
; GENERAL INFORMATION:  
; APPLICANT: Maffitt, Mark A.  
; APPLICANT: Niles, Andrew L.  
; APPLICANT: Haak-Frendscho, Mary  
; TITLE OF INVENTION: Enzymatically-Active Recombinant Human  
; TITLE OF INVENTION: Beta-Trypsin and Method of Making Same  
; NUMBER OF SEQUENCES: 6  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Intellectual Property Department  
; STREET: 8000 Excelsior Drive, Suite 401





NAME: Plumer, Elizabeth R.

TELEFAX: 617-720-2441  
TELEX:

; INFORMATION FOR SEQ ID NO: 4:



```

/ TELEPHONE: 617-720-3500
/ TELEFAX: 617-720-2441
/
/ INFORMATION FOR SEQ ID NO: 2:
/ SEQUENCE CHARACTERISTICS:
/ LENGTH: 2397 base pairs
/ TYPE: nucleic acid
/ STRANDEDNESS: single
/ TOPOLOGY: linear
US-08-978-404B-2

Alignment Scores:
Pred. No.: 0.482 Length: 2397
Score: 86.50 Matches: 29
Percent Similarity: 38.89% Conservative: 20
Best Local Similarity: 23.02% Mismatches: 32
Query Match: 15.67% Indels: 45
DB: 2 Gaps: 5

US-10-020-441-2 (1-106) x US-08-978-404B-2 (1-2397)

QY 7 ArgAnglyAspGlnGlnIlyle-----HisHsGlnProSerGlyValIysVal 23
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
    626 AGCAATGGGTAGAGAGAGTGGGTAGAGATGGATGATCAGACCACTCT----- 673
QY 24 AlaProGlyTyrMet----- 28
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
    674 ---CTGGGTTTCTGGAGAGACAGACAGGCTCTTAATGAATATCTTCTTTAGGAT 730
QY 29 -----ProSerCysMetSerAlaArgGlnArgArg----- 38
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
    731 GTTGCTGACCCCAAGAGTACAGATGAGCTCCGTAAGAGATGCTTATTCATGAC 790
QY 39 -----ProIleHlaGlnThrLeuSer-----Gly 46
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
    791 CACTGATGACTGTGAGCCAGATCATCACACACCCGACTTCAATCGTCAGATGG 850
QY 47 PheAspIleAlaIleValMetLeuAlaGlnMetValAsnLeuGlnSerGlyIleArgVal 66
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
    851 GCAGACATTCCTCTGTAACCTCACAAACCCCTGTGACATTTCTGACATGTCACCT 910
QY 67 IleserLeuProGlnProSerAspIleProProProGlyThrGlyValPheIleValGly 86
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
    911 GTCCCTCACTCTCTGCTGACAGACCTTCCCTCAGAAAGCTGTGCTGGTGGACAGGC 970
QY 87 TyrGlyArgAspAspAsn 92
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
    971 TGGGTATACATGACCAAT 988
DB
RESULT 26
US-08-978-404B-9
/ Sequence 9, Application US/08978404B
/ Patent No. 5968782
/ GENERAL INFORMATION:
/ APPLICANT: Stevens, Richard L.
/ TITLE OF INVENTION: MAST CELL PROTEASE THAT CLEAVES
/ TITLE OF INVENTION: FIBRINOGEN
/ NUMBER OF SEQUENCES: 74
/ CORRESPONDENCE ADDRESS:
/ ADDRESSEE: Wolf, Greenfield & Sacks, P.C.
/ STREET: 600 Atlantic Avenue
/ CITY: Boston
/ STATE: MA
/ COUNTRY: U.S.A.
/ ZIP: 02210-2211
/ COMPUTER READABLE FORM:
/ MEDIUM TYPE: Diskette
/ COMPUTER: IBM Compatible
/ OPERATING SYSTEM: DOS
/ SOFTWARE: FastSeq for Windows Version 2.0
/ CURRENT APPLICATION DATA:
/ APPLICATION NUMBER: US/08/978,404B
/ FILING DATE: 25-NOV-97

```

```

/ CLASSIFICATION: 435
/ PRIOR APPLICATION DATA:
/ APPLICATION NUMBER: 60/032,354
/ FILING DATE: 04-DEC-1996
/ ATTORNEY/AGENT INFORMATION:
/ NAME: Plumer, Elizabeth R.
/ REGISTRATION NUMBER: 36,637
/ REFERENCE/DOCKET NUMBER: B0801/7090
/ TELECOMMUNICATION INFORMATION:
/ TELEPHONE: 617-720-3500
/ TELEFAX: 617-720-2441
/
/ INFORMATION FOR SEQ ID NO: 9:
/ SEQUENCE CHARACTERISTICS:
/ LENGTH: 1095 base pairs
/ TYPE: nucleic acid
/ STRANDEDNESS: single
/ TOPOLOGY: linear
US-08-978-404B-9

Alignment Scores:
Pred. No.: 0.227 Length: 1095
Score: 85.50 Matches: 19
Percent Similarity: 57.14% Conservative: 9
Best Local Similarity: 38.78% Mismatches: 20
Query Match: 15.49% Indels: 1
DB: 2 Gaps: 1

US-10-020-441-2 (1-106) x US-08-978-404B-9 (1-1095)

QY 48 AspIleAlaIleValMetLeuAlaGlnMetValAsnLeuGlnSerGlyIleArgValIle 67
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
    500 GACATGCGCTCTGTAAGTGGAGGCCCCCTGACGCTCTCCAGAGACCTCAACCTGG 559
QY 68 SerLeuProGlnProSerAspIleProProProGlyThrGlyValPheIleValGlyTyr 87
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
    560 TCCCTCCCTCTCTCTCTCTCTGATGTCCTCCCGGGAGTCTATGCTGGTGGACCGGCTGG 619
QY 88 GlyArgAspAspAsnArgAspPro 96
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
    620 GGA---GACATTGCAGACCAACGACCA 643
DB
RESULT 27
US-09-280-116-57
/ Sequence 57, Application US/09280116A
/ Patent No. 6331427
/ GENERAL INFORMATION:
/ APPLICANT: Robison, Keith E.
/ TITLE OF INVENTION: Nucleic Acid Molecules Encoding Human Protease Homologs
/ FILE REFERENCE: 5800-24, 035800/176965
/ CURRENT APPLICATION NUMBER: US/09/280,116A
/ CURRENT FILING DATE: 1999-03-26
/ NUMBER OF SEQ ID NOS: 268
/ SOFTWARE: Patencin Ver. 2.0
/ SEQ ID NO 57
/ LENGTH: 654
/ TYPE: DNA
/ ORGANISM: Homo sapiens
/ FEATURE:
/ OTHER INFORMATION: trypsin-like serine proteases
US-09-280-116-57

Alignment Scores:
Pred. No.: 0.201 Length: 654
Score: 83.50 Matches: 32
Percent Similarity: 46.46% Conservative: 14
Best Local Similarity: 32.32% Mismatches: 44
Query Match: 15.13% Indels: 10
DB: 4 Gaps: 4

US-10-020-441-2 (1-106) x US-09-280-116-57 (1-654)

QY 7 ArgAnglyAspGlnGlnIlyleHisHsGlnProSerGlyValIysVal-----ValAla 24

```



```

Db      259 CGCAACACCTGTGAGACGTCCTGACAGGTCCTGCGGCGAAGCAGCTAGTGACAG 318
Qy      25 ProGlyTYrMetProSerCyMetSerAlaArgGlnArgY-----ProIle 40
Db      319 CCGGGA-----CCACAGCCTATGATGCGCCGGGTGAGCAGGTGAGCAACCCCTG 372
Qy      41 AlaGln---ThrLeuSerGlyPheAspIleAlaIleValMetLeuAlaGlnMetValaen 59
Db      373 TACCAAGGGCAGCGCCTCCAGCGCTGACGTCGCTGTGAGTGGAGGACCAAGTGCCC 432
Qy      60 LeuGlnSerGlyIleArgValIleSerLeuProGlnProSerAspIleProProGly 79
Db      433 TTCACCAATTACATCTCCCTCCGCTGTCCTGACCCCTCGGTGATCTTTGAGACGGGC 492
Qy      80 ThrGlyValPheIleValGlyTYrGlyArgAspAspAspAspAspProSerArg 98
Db      493 ATGAACCTGCTGGGTACTGCTGGGGCAGCCCACTGAGGA-AGACCTCTGCCCCGA 548

RESULT 28
US-09-088-651-6
; Sequence 6, Application US/09088651
; Patent No. 6165771
; GENERAL INFORMATION:
; APPLICANT: BURGESS, NICOLA A.
; APPLICANT: CLINKENBEARD, HELEN E.
; APPLICANT: SOUTHAN, CHRISTOPHER D.
; TITLE OF INVENTION: NOVEL COMPOUNDS
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: RATNER & PRESTIA
; STREET: P.O. BOX 980
; CITY: VALLEY FORGE
; STATE: PA
; COUNTRY: USA
; ZIP: 19482
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FASTSEQ for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/088.651
; FILING DATE: JUNE 1, 1998
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: GB9712088.5
; FILING DATE: 10-JUNE-1997
; APPLICATION NUMBER: EP 97308295.1
; FILING DATE: 17-OCT-1997
; APPLICATION NUMBER: GB 9803650.2
; FILING DATE: 20-FEB-1998
; ATTORNEY/AGENT INFORMATION:
; NAME: PRESTIA, PAUL F.
; REGISTRATION NUMBER: 23,031
; REFERENCE/DOCKET NUMBER: GH30358
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 610-407-0700
; TELEFAX: 610-407-0701
; TELEX: 846169
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1109 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; US-09-088-651-6

Alignment Scores:
Pred. No.: 0.403 Length: 1109
Score: 83.50 Matches: 32
Percent Similarity: 46.46% Mismatches: 14
Conservative: 14

```

```

Best Local Similarity: 32.32% Mismatches: 44
Query Match: 15.13% Indels: 10
DB: 4 Gaps: 4

US-10-020-441-2 (1-106) x US-09-088-651-6 (1-1109)
Qy      7 ArgAsnGlyAspGlnGlnGlyIleHisGlnProSerGlyValIly-----ValAla 24
Db      429 CGCAACACNTCTGAGACGTCCTGTACACAGGTCCTGTGGGGGCAAGGACGTAAGTGACAG 488
Qy      25 ProGlyTYrMetProSerCyMetSerAlaArgGlnArgY-----ProIle 40
Db      489 CCGGGA-----CCACAGCCTATGATGCGCCGGGTGAGCAGGTGAGCAACCCCTG 542
Qy      41 AlaGln---ThrLeuSerGlyPheAspIleAlaIleValMetLeuAlaGlnMetValaen 59
Db      543 TACCAAGGGCAGCGCCTCCAGCGCTGACGTCGCTGTGAGTGGAGGACCAAGTGCCC 602
Qy      60 LeuGlnSerGlyIleArgValIleSerLeuProGlnProSerAspIleProProGly 79
Db      603 TTCACCAATTACATCTCCCTCCGCTGTCCTGACCCCTCGGTGATCTTTGAGACGGGC 662
Qy      80 ThrGlyValPheIleValGlyTYrGlyArgAspAspAspAspProSerArg 98
Db      663 ATGAACCTGCTGGGTACTGCTGGGGCAGCCCACTGAGGA-AGACCTCTGCCCCGA 718

RESULT 29
US-09-386-653A-1
; Sequence 1, Application US/09386653A
; Patent No. 6458564
; GENERAL INFORMATION:
; APPLICANT: Andrade-Gordon, Patricia
; APPLICANT: Darrow, Andrew
; APPLICANT: Qi, Jian-shen
; TITLE OF INVENTION: DNA encoding the novel human serine
; FILE OF INVENTION: ORT-1032
; FILE REFERENCE: ORT-1032
; CURRENT APPLICATION NUMBER: US/09/386.653A
; CURRENT FILING DATE: 1999-08-31
; NUMBER OF SEQ ID NOS: 11
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
; LENGTH: 1110
; TYPE: DNA
; ORGANISM: Homo sapiens
; US-09-386-653A-1

Alignment Scores:
Pred. No.: 0.403 Length: 1110
Score: 83.50 Matches: 32
Percent Similarity: 46.46% Mismatches: 14
Best Local Similarity: 32.32% Mismatches: 44
Query Match: 15.13% Indels: 10
DB: 4 Gaps: 4

US-10-020-441-2 (1-106) x US-09-386-653A-1 (1-1110)
Qy      7 ArgAsnGlyAspGlnGlnGlyIleHisGlnProSerGlyValIly-----ValAla 24
Db      267 CGCAACACCTGTGAGACGTCCTGTACACAGGTCCTGTGGGGGCAAGGACGTAAGTGAG 326
Qy      25 ProGlyTYrMetProSerCyMetSerAlaArgGlnArgY-----ProIle 40
Db      327 CCGGGA-----CCACAGCCTATGATGCGCCGGGTGAGCAGGTGAGCAACCCCTG 380
Qy      41 AlaGln---ThrLeuSerGlyPheAspIleAlaIleValMetLeuAlaGlnMetValaen 59
Db      381 TACCAAGGGCAGCGCCTCCAGCGCTGACGTCGCTGTGAGTGGAGGACCAAGTGCCC 440
Qy      60 LeuGlnSerGlyIleArgValIleSerLeuProGlnProSerAspIleProProGly 79
Db      441 TTCACCAATTACATCTCCCTCCGCTGTCCTGACCCCTCGGTGATCTTTGAGACGGGC 500

```

Qy 80 ThrGlyValPheIleValGlyTyrGlyArgAspAsnAspArgAspProSerArg 98  
Db 501 ATGAAGCTGCTGGGTACATCGGCTGGGGCAGCCCACTGAGGA-AGACCTCTCTCCCGA 556  
RESULT 30  
US-09-386-653A-8  
Sequence 8, Application US/09386653A  
Patent No. 6458564  
GENERAL INFORMATION:  
APPLICANT: Andrade-Gordon, Patricia  
APPLICANT: Darrow, Andrew  
TITLE OF INVENTION: DNA encoding the novel human serine  
PROTEASE T  
FILE REFERENCE: ORT-1032  
CURRENT APPLICATION NUMBER: US/09/386,653A  
CURRENT FILING DATE: 1999-08-31  
NUMBER OF SEQ ID NOS: 11  
SOFTWARE: PatentIn Ver. 2.0  
SEQ ID NO 8  
LENGTH: 1130  
TYPE: DNA  
ORGANISM: Artificial Sequence  
FEATURE:  
OTHER INFORMATION: Description of Artificial Sequence: Fusion gene of  
US-09-386-653A-8  
Alignment Scores:  
Pred. No.: 0.413 Length: 1130  
Score: 83.50 Matches: 32  
Percent Similarity: 46.46% Conservative: 14  
Best Local Similarity: 32.32% Mismatches: 44  
Query Match: 15.13% Indels: 10  
DB: 4 Gaps: 4  
US-10-020-441-2 (1-106) x US-09-386-653A-8 (1-1130)  
Qy 7 ArgAsnGlyAspGlyGlnGlyIleHisGlnProSerGlyValIleVal-ValAla 24  
Db 235 CCGAACACCTCTGAGACGCTCCCTGACAGGTCCTGCTGGGGCAGAGCACTAGTGCG 354  
Qy 25 ProGlyTyrMetProSerCysMetSerAlaArgGlnArg-Protile 40  
Db 355 CCGGGA-----CCACAGCTATGTATGCCCGGGTGAGGAGTGAGCAACCCCTG 408  
Qy 41 AlaGln---ThrLeuSerGlyPheAspIleAlaIleValMetLeuAlaGlnMetValAsn 59  
Db 409 TACCAAGGCGACGCGCTCCAGCGCTGACGTGGCCCTGTGAGCTGAGGACACCATGCCCC 468  
Qy 60 LeuGlnSerGlyIleArgValIleSerLeuProGlnProSerAspIleProProGly 79  
Db 469 TTCACCAATTACATCTCTCCCGCGTGCTGCTGACCCCTCGGTGATCTTTGAGACGGCC 528  
Qy 80 ThrGlyValPheIleValGlyTyrGlyArgAspAsnAspArgAspProSerArg 98  
Db 529 ATGAAGCTGCTGGGTACATCGGCTGGGGCAGCCCACTGAGGA-AGACCTCTCTCCCGA 584  
RESULT 31  
US-08-978-404B-1  
Sequence 1, Application US/08978404B  
Patent No. 5968782  
GENERAL INFORMATION:  
APPLICANT: Stevens, Richard L.  
TITLE OF INVENTION: MAST CELL PROTEASE THAT CLEAVES  
FIBRINOGEN  
NUMBER OF SEQUENCES: 74  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Wolf, Greenfield & Sacke, P.C.  
STREET: 600 Atlantic Avenue  
CITY: Boston  
STATE: MA  
COUNTRY: U.S.A.

ZIP: 02210-2211  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: DOS  
SOFTWARE: FastSeq for Windows Version 2.0  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/978,404B  
FILING DATE: 25-NOV-97  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 60/032,354  
FILING DATE: 04-DEC-1996  
ATTORNEY/AGENT INFORMATION:  
NAME: Plumer, Elizabeth R.  
REGISTRATION NUMBER: 36,637  
REFERENCE/DOCKET NUMBER: B0801/7090  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 617-720-3500  
TELEFAX: 617-720-2441  
TELEX:  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1031 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
US-08-978-404B-1  
Alignment Scores:  
Pred. No.: 0.42 Length: 1031  
Score: 83.00 Matches: 17  
Percent Similarity: 55.32% Conservative: 9  
Best Local Similarity: 36.17% Mismatches: 21  
Query Match: 15.04% Indels: 0  
DB: 2 Gaps: 0  
US-10-020-441-2 (1-106) x US-08-978-404B-1 (1-1031)  
Qy 46 GlyPheAspIleAlaIleValMetLeuAlaGlnMetValAsnLeuGlnSerGlyIleArg 65  
Db 375 GGGGCGACATGTCCTGCTGTAACATCAGAACACCTGTGAACTTTGCATATGTCCAC 434  
Qy 66 ValIleSerLeuProGlnProSerAspIleProProGlyThrGlyValPheIleVal 85  
Db 435 CCTGTCCCTTACTCTCCGCTCAGAGACCTTCCCTCAGGAACGTGTGCTGGGTGACA 494  
Qy 86 GlyTyrGlyArgAspAsn 92  
Db 495 GGCTGGGTATACATGACAT 515  
RESULT 32  
US-09-027-337-1  
Sequence 1, Application US/09027337B  
Patent No. 5972616  
GENERAL INFORMATION:  
APPLICANT: O'Brien, Timothy J.  
APPLICANT: Tanimoto, Hiroshi  
TITLE OF INVENTION: TADG-15: An Extracellular Serine Protease Overexpressed in  
BREAST AND OVARIAN CARCINOMAS  
FILE REFERENCE: D6064  
CURRENT APPLICATION NUMBER: US/09/027,337B  
CURRENT FILING DATE: 1998-02-20  
NUMBER OF SEQ ID NOS: 13  
SEQ ID NO 1  
LENGTH: 3147  
TYPE: DNA  
ORGANISM: Homo sapiens  
FEATURE:  
LOCATION: 23..2589  
OTHER INFORMATION: cDNA sequence of TADG-15  
US-09-027-337-1

Alignment Scores:

Pred. No.:	Score:	Percent Similarity:	Best Local Similarity:	Query Match:	Length:	Matches:	Conservative:	Mismatches:	Indels:	Gaps:
2.1	82.50	44.71%	27.06%	14.95%	3147	23	15	28	19	3

US-10-020-441-2 (1-106) x US-09-027-337-1 (1-3147)

Qy 13 GYLIEHSHISGLINProSerGIYValIysValAlaProGIYTYrMeProSerCysMet 32  
 Db 2051 GGCTTGACGACGACGAGCCAG-----CGCAGCGCCCTGGGGTG----- 2089

Qy 33 SerAlaArgGlnArgProIleAlaGlnThrLeuSer----- 45  
 Db 2090 -----GAGGAGCGGAGGCTCAAGCCGATCATCTCCACCCCTTTCATGACTTC 2140

Qy 46 -----GlyPheAspIleAlaIleValMetLeuAlaGlnMetValAsnLeuGlnSerGly 63  
 Db 2141 ACCTTCGACTATGACATCGCGCTGCTGGAGCTGGAGAAACCGGACAGATACAGCTCCATG 2200

Qy 64 IleArgValIleSerLeuProGlnProSerAspIleProProProGIYThrGlyValPhe 83  
 Db 2201 GTGCGGCCCATCTGCGCTCGGAGCGCTCCATGCTTCCTGCGCGGAGGCCATCTGG 2260

Qy 84 IleValGIYTYrGIY 88  
 Db 2261 GTCAAGGGCTGGGGA 2275

RESULT 33  
 US-09-644-600-1  
 ; Sequence 1, Application US/09644600  
 ; Patent No. 6451500  
 ; GENERAL INFORMATION:  
 ; APPLICANT: O'Brien, Timothy J.  
 ; APPLICANT: Tanimoto, Hirotooshi  
 ; TITLE OF INVENTION: TADG-15: An Extracellular Serine Protease  
 ; FILE REFERENCE: D6064CIP/D  
 ; CURRENT APPLICATION NUMBER: US/09/644,600  
 ; CURRENT FILING DATE: 2000-08-23  
 ; PRIOR APPLICATION NUMBER: 09/421,213  
 ; PRIOR FILING DATE: 1999-10-20  
 ; PRIOR APPLICATION NUMBER: 09/027,337  
 ; NUMBER OF SEQ ID NOS: 98  
 ; SEQ ID NO 1  
 ; LENGTH: 3147  
 ; TYPE: DNA  
 ; ORGANISM: Homo sapiens  
 ; FEATURE:  
 ; OTHER INFORMATION: TADG-15  
 US-09-644-600-1

Alignment Scores:

Pred. No.:	Score:	Percent Similarity:	Best Local Similarity:	Query Match:	Length:	Matches:	Conservative:	Mismatches:	Indels:	Gaps:
2.1	82.50	44.71%	27.06%	14.95%	3147	23	15	28	19	3

US-10-020-441-2 (1-106) x US-09-644-600-1 (1-3147)

Qy 13 GYLIEHSHISGLINProSerGIYValIysValAlaProGIYTYrMeProSerCysMet 32  
 Db 2051 GGCTTGACGACGACGAGCCAG-----CGCAGCGCCCTGGGGTG----- 2089

Qy 33 SerAlaArgGlnArgProIleAlaGlnThrLeuSer----- 45  
 Db 2090 -----GAGGAGCGGAGGCTCAAGCCGATCATCTCCACCCCTTTCATGACTTC 2140

Qy 46 -----GlyPheAspIleAlaIleValMetLeuAlaGlnMetValAsnLeuGlnSerGly 63  
 Db 2141 ACCTTCGACTATGACATCGCGCTGCTGGAGCTGGAGAAACCGGACAGATACAGCTCCATG 2200

Qy 64 IleArgValIleSerLeuProGlnProSerAspIleProProProGIYThrGlyValPhe 83  
 Db 2201 GTGCGGCCCATCTGCGCTCGGAGCGCTCCATGCTTCCTGCGCGGAGGCCATCTGG 2260

Qy 84 IleValGIYTYrGIY 88  
 Db 2261 GTCAAGGGCTGGGGA 2275

RESULT 34  
 US-09-644-600-18/c  
 ; Sequence 18, Application US/09644600  
 ; Patent No. 6451500  
 ; GENERAL INFORMATION:  
 ; APPLICANT: O'Brien, Timothy J.  
 ; APPLICANT: Tanimoto, Hirotooshi  
 ; TITLE OF INVENTION: TADG-15: An Extracellular Serine Protease  
 ; FILE REFERENCE: D6064CIP/D  
 ; CURRENT APPLICATION NUMBER: US/09/644,600  
 ; CURRENT FILING DATE: 2000-08-23  
 ; PRIOR APPLICATION NUMBER: 09/421,213  
 ; PRIOR FILING DATE: 1999-10-20  
 ; PRIOR APPLICATION NUMBER: 09/027,337  
 ; NUMBER OF SEQ ID NOS: 98  
 ; SEQ ID NO 18  
 ; LENGTH: 3147  
 ; TYPE: RNA  
 ; ORGANISM: Artificial sequence  
 ; FEATURE:  
 ; OTHER INFORMATION: Antisense of TADG-15  
 US-09-644-600-18

Alignment Scores:

Pred. No.:	Score:	Percent Similarity:	Best Local Similarity:	Query Match:	Length:	Matches:	Conservative:	Mismatches:	Indels:	Gaps:
2.1	82.50	44.71%	27.06%	14.95%	3147	23	15	28	19	3

US-10-020-441-2 (1-106) x US-09-644-600-18 (1-3147)

Qy 13 GYLIEHSHISGLINProSerGIYValIysValAlaProGIYTYrMeProSerCysMet 32  
 Db 1097 GGCTTGACGACGACGAGCCAG-----CGCAGCGCCCTGGGGTG----- 1059

Qy 33 SerAlaArgGlnArgProIleAlaGlnThrLeuSer----- 45  
 Db 1058 -----GAGGAGCGGAGGCTCAAGCCGATCATCTCCACCCCTTTCATGACTTC 1008

Qy 46 -----GlyPheAspIleAlaIleValMetLeuAlaGlnMetValAsnLeuGlnSerGly 63  
 Db 1007 ACCTTCGACTATGACATCGCGCTGCTGGAGCTGGAGAAACCGGACAGATACAGCTCCATG 948

Qy 64 IleArgValIleSerLeuProGlnProSerAspIleProProProGIYThrGlyValPhe 83  
 Db 947 GTGCGGCCCATCTGCGCTCGGAGCGCTCCATGCTTCCTGCGCGGAGGCCATCTGG 888

Qy 84 IleValGIYTYrGIY 88  
 Db 887 GTCAAGGGCTGGGGA 873

RESULT 35  
 US-08-681-151-2  
 ; Sequence 2, Application US/08681151  
 ; Patent No. 5869637  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Au-Young, Janice

```

APPLICANT: Bandman, Olga
APPLICANT: Braxton, Scott Michael
APPLICANT: Goli, Surya
TITLE OF INVENTION: A NOVEL HUMAN KALLIKREIN
NUMBER OF SEQUENCES: 4
CORRESPONDENCE ADDRESSES:
ADDRESS: INCYTE PHARMACEUTICALS, INC.
STREET: 3174 Porter Drive
CITY: Palo Alto
STATE: CA
COUNTRY: US
ZIP: 94304
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSeq Version 1.5
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/681,151
FILING DATE: Herewith
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Billings, Lucy J.
REGISTRATION NUMBER: 36,749
REFERENCE/DOCKET NUMBER: PF-0074US
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415-855-0555
TELEFAX: 415-845-4166
TELEX:
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 1739 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cdna
IMMEDIATE SOURCE:
LIBRARY: HEARNOT01
CLONE: 307474
US-08-681-151-2

Alignment Scores:
Pred. No.: 1.46 Length: 1739
Score: 81.00 Matches: 22
Percent Similarity: 44.58% Conservative: 15
Best Local Similarity: 26.51% Mismatches: 45
Query Match: 14.67% Indels: 1
DB: 2 Gaps: 0

US-10-020-441-2 (1-106) x US-08-681-151-2 (1-1739)
QY 11 gmgmglylshshsglnproserglyvallyvalalaproglytyrmetproser 30
DB 1105 GAAAGTGGTGGCTGATCAACATCTAGACCATCATCAGTTCATGACAGCAGCTT 1164
QY 31 CysMetSerAlaArgGlnArgArgProIleAlaGlnThrLeuSerGly-PheAspIleAl 50
DB 1165 TGTGAAGACCATCATCTCCATCCCGCTACAGTCGACGAGTGGTGGTACTATGACATCAG 1224
QY 50 allevaImetleuAlaGlnMetValasmeuGlnserGlyIleArgValIIsersleuPr 70
DB 1225 CATCGTGGCTGAGGAAAGACATCAAGTAGAGAGCTGCTGCTGGCTGTGCTTGCC 1284
QY 70 oGlnProSerAspIleProProProGlyThrGlyValPheIleValGlyTyrgIlyArgAs 90
DB 1289 CAACCGGAGCAGTGGCTAGAGCTGACACGTACTGTATATACAGGCTGGGGCCACAT 1344
QY 90 pAspAsn 92
DB 1345 GGGCAAT 1351

```

```

RESULT 36
US-08-666-082B-2
; Sequence 2, Application US/08666082B
; Patent No. 5916770
; GENERAL INFORMATION:
APPLICANT: YOSHIKAWA, WATARU
APPLICANT: SHIMONISHI, MANABU
APPLICANT: IWAMOTO, JUNKO
APPLICANT: TAKEHARA, TOYOHIRO
APPLICANT: HAGIYA, MICHIO
TITLE OF INVENTION: MACROPHAGE STIMULATING PROTEIN VARIANT AND
TITLE OF INVENTION: METHOD FOR PRODUCING THE SAME
NUMBER OF SEQUENCES: 7
CORRESPONDENCE ADDRESSES:
ADDRESSEE: LEYDIG, VOIT & MAYER, LTD.
STREET: TWO PRUDENTIAL PLAZA, SUITE 4900
CITY: CHICAGO
STATE: IL
COUNTRY: US
ZIP: 60601-6780
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/666,082B
FILING DATE: 19-JUN-1996
CLASSIFICATION: 536
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 7-153309
FILING DATE: 20-JUN-1995
ATTORNEY/AGENT INFORMATION:
NAME: ROBERT F. GREEN
REGISTRATION NUMBER: 27555
REFERENCE/DOCKET NUMBER: 73843
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 2216 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: cdna
ORIGINAL SOURCE:
ORGANISM: Human
FEATURE:
NAME/KEY: sig peptide
LOCATION: 1..54
FEATURE:
NAME/KEY: CDS
LOCATION: 1..2133
US-08-666-082B-2

Alignment Scores:
Pred. No.: 2.65 Length: 2216
Score: 80.00 Matches: 32
Percent Similarity: 41.18% Conservative: 17
Best Local Similarity: 26.89% Mismatches: 30
Query Match: 14.49% Indels: 40
DB: 2 Gaps: 6

US-10-020-441-2 (1-106) x US-08-666-082B-2 (1-2216)
QY 4 LeuThrIleuArgAsArgIAspGlnGlnGlyIleHis----- 15
DB 1486 GTACACTTGGCGGAAT-----CGGACAGGCGCAGATTTCCTGGGGGGGTCTCTAGTGAAG 1539
QY 16 -----HisGlnPro-----serGly 20
DB 1540 GAGCAGTGGATCTAGACTGCCCCGGCAGTGTCTCTCTGCAATATAGCTCTTACAGGGC 1599
QY 21 ValIysValAlaIProGlyTyrimet-----ProSerCysMet 32

```

```

Db 1600 TATGAGGTATGTTGGCACCCTGTTCAGAACCCACAGATGAGAGCCAAAGCTTA--- 1656
Oy 33 SerAlaValArgGlnArgArgProIleAlaGlnThrIleu-----SerGlyPheAspIle 49
Db 1657 -----CAGCGGATCCCAAGTACCAAGATGATGTGTGAGCCCTCAGAGCTTCCCAAGCTT 1707
Oy 50 AlaIleValMetLeuAlaGlnMetValAsnLeuGlnSerGlyIleArgValIleSerIleu 69
Db 1708 GTTCCCTGACCACTGACAGAGATCTGTGACCCCTGAACCAAGCGGTGGCCCTCAGATTCGAGCTTG 1767
Oy 70 ProGlnProSerArgIleProProProGlyThrGlyValPheIleValGlyTyrGly 88
Db 1768 CCCCTGAAATGATATGTGTGCTGCTTCCAGGAGCAACAGTGTGAGATTGCAGAGCTGGAGT 1824

RESULT 37
US-07-882-925A-1
: Sequence 1, Application US/07882925A
: Patent No. 5315000
:
: GENERAL INFORMATION:
: APPLICANT: Degen, Sandra J. F.
: TITLE OF INVENTION: Gene for a growth factor and its cDNA and
: TITLE OF INVENTION: protein
: NUMBER OF SEQUENCES: 7
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Gregory Lunn
: STREET: Wood, Herron & Evans, 2700 Carew Tower
: CITY: Cincinnati
: STATE: Ohio
: COUNTRY: USA
: ZIP: 45202
:
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Diskette, 3.50 inch, 800 Kb
: COMPUTER: Apple Macintosh
: OPERATING SYSTEM: Macintosh 6.0.3
: SOFTWARE: Microsoft Word 4.0
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/07/882,925A
: FILING DATE: 19920514
: CLASSIFICATION: 530
: ATTORNEY/AGENT INFORMATION:
: NAME: Lunn, Gregory
: REGISTRATION NUMBER: 29,945
: REFERENCE/DOCKET NUMBER: CMC 57
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (513) 241-2324
: TELEFAX: (513) 421-7269
: INFORMATION FOR SEQ ID NO: 1:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 2219 base pairs
: TYPE: NUCLEIC ACID
: STRANDEDNESS: single
: TOPOLOGY: linear
: MOLECULE TYPE: cDNA to mRNA
: ANTI-SENSE: no
: ORIGINAL SOURCE:
: ORGANISM: human
: DEVELOPMENTAL STAGE: fetal
: TISSUE TYPE: liver
: IMMEDIATE SOURCE:
: LIBRARY: cDNA
: CLONE: #33
: POSITION IN GENOME:
: CHROMOSOME/SEGMENT: human 3p21/D3P1552
: FEATURE:
: IDENTIFICATION METHOD: experimental
: OTHER INFORMATION: Includes five polymorphisms at the nucleotide
: OTHER INFORMATION: level; one of which results in an amino acid substitution (nu-
: OTHER INFORMATION: 619). Sequence ID NO:2: contains the identical sequence with
: PUBLICATION INFORMATION:
: RELEVANT RESIDUES IN SEQ ID NO: 1: FROM 1 TO 2219
US-07-882-925A-1

```

Alignment Scores:	2.66	length:	2219
Pred. No.:	80.00	Matches:	32
Score:	41.188	Conservative:	17
Percent Similarity:	26.894	Mismatches:	30
Best Local Similarity:	14.494	Indels:	40
Query Match:	1	Gaps:	6
DB:			

US-10-020-441-2 (1-106) x US-07-882-925A-1 (1-2219)

Qy	4	LeuThrIeuArgaEngIyaSpGngIngIyIleHis	-----	15
Db	1470	GTCAGCTTGGCCGAAT	-----CGGACAGGCCAGACATTTCTGCGGGGGGCTCTAGTGAAG	15233
Qy	16	-----	-----HicGlnPro-----SerIy 20	
Db	1524	GAGCAGTGATACGACTGCCGCGCAGTGTCTCTCTCTGCAATATGCTTCAAGG3C	15833	
Qy	21	ValIyValAlaIaPProGlyIyTmet	-----ProSerCySmet 32	
Db	1584	TATAGAGTATGCTGGGACACCTCTTCCAGAACCCACAGCATGAGAGCCAGCTTA	16410	
Qy	33	SerIaIaArgIaArgaPProIleIaIaGlnThrIeu	-----SerGlyPheAspIle 49	
Db	1641	-----CAGCGCGGCGCCAGACGACAGATGATGATGCGCCCTCAGGCTCCAGCTT	16919	
Qy	50	AlaIleValMetIeuIaGlnMetValaIenIeuGlnSerGlyIleArgValIleSerIeu	69	
Db	1692	GTCCTGCTCAGCTGAGAGATCTGTGACCTGTGAACCAAGCGTGG3CCTGATTCGCTG	17511	
Qy	70	ProGlnProSerAspIlePProPProGlyIyThrGlyValaIleValaIyIyGly 88		
Db	1752	CCCCCTGAATGATATGTGTGCTCTCCAGGAGCCACATGTGAATTCAGAGCTGGGT	1808	

RESULT 38  
US-07-882-925A-2  
Sequence 2, Application US/07882925A  
Patent No. 5315000  
GENERAL INFORMATION:  
APPLICANT: Degen, Sandra J. F.  
TITLE OF INVENTION: Gene for a growth factor and its cDNA and  
TITLE OF INVENTION: protein  
NUMBER OF SEQUENCES: 7  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Gregory Lunn  
STREET: Wood, Herion & Evans, 2700 Carew Tower  
CITY: Cincinnati  
STATE: Ohio  
COUNTRY: USA  
ZIP: 45202  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette, 3.50 inch, 800 Kb  
COMPUTER: Apple Macintosh  
OPERATING SYSTEM: Macintosh 6.0.3  
SOFTWARE: Microsoft Word 4.0  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/07/882,925A  
FILING DATE: 19920514  
CLASSIFICATION: 530  
ATTORNEY/AGENT INFORMATION:  
NAME: Lunn, Gregory  
REGISTRATION NUMBER: 29,945  
REFERENCE/DOCKET NUMBER: CMC 57  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (513) 241-7269  
TELEFAX: (513) 421-7324  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 2219 base pairs  
TYPE: NUCLEIC ACID  
STRANDEDNESS: single  
TOPOLOGY: linear

MOLECULE TYPE: cDNA to mRNA  
ANTI-SENSE: no  
ORIGINAL SOURCE:  
ORGANISM: human  
DEVELOPMENTAL STAGE: fetal  
TISSUE TYPE: liver  
IMMEDIATE SOURCE:  
LIBRARY: cDNA  
CLONE: #33  
POSITION IN GENOME:  
CHROMOSOME/SEGMENT: human 3p21/D3F15S2  
FEATURE:  
IDENTIFICATION METHOD: experimental  
OTHER INFORMATION: Includes five polymorphisms at the nucleotide  
OTHER INFORMATION: level: one of which results in an amino acid substitution (nu  
OTHER INFORMATION: 619). Sequence ID NO:1: contains the identical sequence with  
OTHER INFORMATION: polymorphic amino acid.  
PUBLICATION INFORMATION:  
RELEVANT RESIDUES IN SEQ ID NO: 2: FROM 1 TO 2219  
US-07-882-925A-2

Alignment Scores:  
Pred. No.: 2.66 Length: 2219  
Score: 80.00 Matches: 32  
Percent Similarity: 41.18% Conservative: 17  
Best Local Similarity: 26.89% Mismatches: 30  
Query Match: 14.49% Indels: 40  
DB: 1 Gaps: 6

US-10-020-441-2 (1-106) x US-07-882-925A-2 (1-2219)

QY 4 LeuThrLeuArgAsnGlyAaspGlnGlnGlyIleHis-----SerglyPheasp1le 15  
DB 1470 GTCACCTTCCGAAT-----CGGCAAGGCCAGCATTTCTGGGGGGCTCTAGTAAG 1523  
QY 16 -----HieGlnPro---Sergly 20  
DB 1524 GAGCAGTGATCTGACTGCCCCGAGTCTTCTCTCTGCAATGCTCTCAAGCGGC 1583  
QY 21 ValLysValAlaProGlyTyrMet-----ProserCysmet 32  
DB 1584 TATGAGGTATGTTGGTGGCACCTGTTCCAGAACCCACAGCATGAGAGCGCAAGCCTA--- 1640  
QY 33 SerAlaArgGlnArgProIleAlaGlnThrLeu-----SerglyPheasp1le 49  
DB 1641 -----CACCGGTCTCCAGTACGCAAGATGTGTGGGCCCTCAGGCTCCAGCTT 1691  
QY 50 AlaIleValMetLeuAlaGlnMetValAsnLeuGlnSerglyIleArgValIleSeriLeu 69  
DB 1692 GTCCTGCTCAAGCTGAGATCTGTGACCTGAAACACAGCGTGGCCCTGATCTGCTG 1751  
QY 70 ProGlnProSerAspIleProProGlyThrGlyValPheIleValGlyTyrGly 88  
DB 1752 CCCCCTGAATGTATGTGTGTGCTTCCAGGAGCCAAAGTGTGAGATTCAGAGCTGGGGT 1808

RESULT 39  
US-08-184-012C-1  
Sequence 1, Application US/08184012C  
Patent No. 5606029  
GENERAL INFORMATION:  
APPLICANT: Degen, Sandra J. F.  
TITLE OF INVENTION: Gene for a growth factor and its cDNA and  
NUMBER OF SEQUENCES: 10  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Gregory Lunn  
STREET: Wood, Heron & Evans, 2700 Carew Tower  
CITY: Cincinnati  
STATE: Ohio  
COUNTRY: USA  
ZIP: 45202  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette, 3.50 inch, 800 KB

COMPUTER: Apple Macintosh  
OPERATING SYSTEM: Macintosh 7.5.2  
SOFTWARE: Microsoft Word 5.1  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/184,012C  
FILING DATE: 1/18/94  
CLASSIFICATION: 536  
ATTORNEY/AGENT INFORMATION:  
NAME: Lunn, Gregory  
REGISTRATION NUMBER: 29,945  
REFERENCE/DOCKET NUMBER: CMC 57  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (513) 241-2324  
TELEFAX: (513) 421-7269  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 2219 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: cDNA to mRNA  
ANTI-SENSE: no  
ORIGINAL SOURCE:  
ORGANISM: human  
DEVELOPMENTAL STAGE: fetal  
TISSUE TYPE: liver  
IMMEDIATE SOURCE:  
LIBRARY: cDNA  
CLONE: #33  
POSITION IN GENOME:  
CHROMOSOME/SEGMENT: human 3p21/D3F15S2  
FEATURE:  
IDENTIFICATION METHOD: experimental  
OTHER INFORMATION: Includes five polymorphisms at the nucleotide  
OTHER INFORMATION: level: one of which results in an amino acid substitution (nu  
OTHER INFORMATION: 619). Sequence ID NO:2: contains the identical sequence with  
PUBLICATION INFORMATION:  
RELEVANT RESIDUES IN SEQ ID NO: 1: FROM 1 TO 2219  
US-08-184-012C-1

Alignment Scores:  
Pred. No.: 2.66 Length: 2219  
Score: 80.00 Matches: 32  
Percent Similarity: 41.18% Conservative: 17  
Best Local Similarity: 26.89% Mismatches: 30  
Query Match: 14.49% Indels: 40  
DB: 1 Gaps: 6

US-10-020-441-2 (1-106) x US-08-184-012C-1 (1-2219)

QY 4 LeuThrLeuArgAsnGlyAaspGlnGlnGlyIleHis-----SerglyPheasp1le 15  
DB 1470 GTCACCTTCCGAAT-----CGGCAAGGCCAGCATTTCTGGGGGGCTCTAGTAAG 1523  
QY 16 -----HieGlnPro---Sergly 20  
DB 1524 GAGCAGTGATCTGACTGCCCCGAGTCTTCTCTCTGCAATGCTCTCAAGCGGC 1583  
QY 21 ValLysValAlaProGlyTyrMet-----ProserCysmet 32  
DB 1584 TATGAGGTATGTTGGTGGCACCTGTTCCAGAACCCACAGCATGAGAGCGCAAGCCTA--- 1640  
QY 33 SerAlaArgGlnArgProIleAlaGlnThrLeu-----SerglyPheasp1le 49  
DB 1641 -----CACCGGTCTCCAGTACGCAAGATGTGTGGGCCCTCAGGCTCCAGCTT 1691  
QY 50 AlaIleValMetLeuAlaGlnMetValAsnLeuGlnSerglyIleArgValIleSeriLeu 69  
DB 1692 GTCCTGCTCAAGCTGAGATCTGTGACCTGAAACACAGCGTGGCCCTGATCTGCTG 1751  
QY 70 ProGlnProSerAspIleProProGlyThrGlyValPheIleValGlyTyrGly 88

Db 1752 CCCCCTGATGTATGTGTGCTCCAGGACCAAGTGTGATTTGCAGGCTGGGCT 1808  
RESULT 40  
US-08-184-012C-2  
Sequence 2, Application US/08184012C  
Patent No. 5606029  
GENERAL INFORMATION:  
APPLICANT: Degen, Sandra J. F.  
TITLE OF INVENTION: Gene for a growth factor and its cDNA and  
TITLE OF INVENTION: Protein  
NUMBER OF SEQUENCES: 10  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Gregory Lunn  
STREET: Wood, Herron & Evans, 2700 Carew Tower  
CITY: Cincinnati  
STATE: Ohio  
COUNTRY: USA  
ZIP: 45202  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette, 3.50 inch, 800 KB  
COMPUTER: Apple Macintosh  
OPERATING SYSTEM: Macintosh 7.5.2  
SOFTWARE: Microsoft Word 5.1  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/184,012C  
FILING DATE: 1/18/94  
CLASSIFICATION: 536  
ATTORNEY/AGENT INFORMATION:  
NAME: Lunn, Gregory  
REGISTRATION NUMBER: 29,945  
REFERENCE/DOCKET NUMBER: CMC 57  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (513) 241-2324  
TELEFAX: (513) 421-7269  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 2219 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: cDNA to mRNA  
ANTI-SENSE: no  
ORIGINAL SOURCE:  
ORGANISM: human  
DEVELOPMENTAL STAGE: fetal  
TISSUE TYPE: liver  
IMMEDIATE SOURCE:  
LIBRARY: cDNA  
CLONE: #33  
POSITION IN GENOME:  
CHROMOSOME/SEGMENT: human 3p21/D3F1S52  
FEATURE:  
IDENTIFICATION METHOD: experimental  
OTHER INFORMATION: Includes five polymorphisms at the nucleotide  
OTHER INFORMATION: level; one of which results in an amino acid substitution (m  
OTHER INFORMATION: 619). Sequence ID NO:1: contains the identical sequence with  
Patent No. 5606029  
OTHER INFORMATION: polymorphic amino acid.  
PUBLICATION INFORMATION:  
RELEVANT RESIDUES IN SEQ ID NO: 2: FROM 1 TO 2219  
US-08-184-012C-2  
Alignment Scores:  
Pred. No.: 2.66 Length: 2219  
Score: 80.00 Matches: 32  
Percent Similarity: 41.18% Conservative: 17  
Best Local Similarity: 26.89% Mismatches: 30  
Query Match: 14.49% Indels: 40  
DB: 1 Gaps: 6  
US-10-020-441-2 (1-106) x US-08-184-012C-2 (1-2219)

Qy 4 LeuThrLeuArgAsnGlyAspDlnGlnGlyIleHis----- 15  
Db 1470 GTACAGCTTGCCGAAT-----CGGACGGGCCAGCATTTCTCGGGGGGGCTCTTAGTAGAG 1523  
Qy 16 -----HisGlnPro---SerGly 20  
Db 1524 GAGCAGTGAATAGTACTGCCCCGGCAGTCTTCTCTCTCCATATGCTCTACAGGGC 1583  
Qy 21 VallysValAlaProGlyTyrMet-----ProSerCysMet 32  
Db 1584 TATGAGTATGTTGGGACCCCTGTCCAGAACCCACAGCATGAGAGCCAGCCTA---- 1640  
Qy 33 SerAlaArgIleArgArgProIleAlaGlnThrIleu-----SerGlyPheAspIle 49  
Db 1641 -----CAGCGGGTCCAGTAGCCAAAGATGTTGTGGGCCCTCAGGGCTCCAGCTT 1691  
Qy 50 AlaIleValMetLeuAlaGlnMetValAsnLeuGlnSerGlyIleArgValIleSerIle 69  
Db 1692 GTCCGTCAAGCTGAGAGATCTGTGACCTGAACCGCGGTGGCCCTGATCTGCCTG 1751  
Qy 70 ProGlnProSerAspIleProProProGlyThrGlyValPheIleValGlyTyrGly 88  
Db 1752 CCCCCTGATGTATGTGTGCTCCAGGACCAAGTGTGATTTGCAGGCTGGGCT 1808

Search completed: April 8, 2003, 19:53:28  
Job time : 63.8689 secs

**THIS PAGE BLANK (USPTO)**



GenCore version 5.1.4 ps 4578  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: April 8, 2003, 10:21:07 ; Search time 212 Seconds  
(without alignments)  
1125.999 Million cell updates/sec

Title: US-10-020-441-2

Perfect score: 552  
Sequence: 1 VSFLTRNGDQGIHHQPSG.....YGRDNDNRDPKRGKGIILKK 106

Scoring table: BLOSUM62  
Xgapop 10.0, Xgapext 0.5  
Ygapop 10.0, Ygapext 0.5  
Fgapop 6.0, Fgapext 7.0  
Delop 6.0, Delext 7.0

Searched: 2185239 seqs, 1125999159 residues

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Command line parameters:  
MODEL=frame+ p2n model -DEV=xlh  
Q=/sgn2.1/USPTO/spool/US10020441/runat\_02042003\_093011\_20959/app\_query.fasta\_1.462  
-DB=N Geneseq\_101002 -QFMT=fastap -SUFIX=ring -MINMATCH=0.1 -LOOPCL=0  
-LOOPEXT=0 -UNIT5=bits -START=1 -END=-1 -MATRIX=biosum62 -TRANS=human40.cdi  
-LIST=45 -DOCCALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=40  
-MODE=LOCAL -OUTFMT=pcio -NORF=ext -HEADSIZE=500 -MINLEN=0 -MAXLEN=2000000000  
-USER=US10020441 @GCG 1.1 275 @runat\_02042003\_093011\_20959 -NCPU=6 -ICPU=3  
-NO\_XLPRY -NO\_MMAP -LARGEQUERY -NEG\_SCORES=0 -WAIT -LONGLOG -DEV\_TIMEOUT=120  
-WARN\_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -Fgapop=6 -Fgapext=7  
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : N Geneseq\_101002.\*  
1: /SID2/gcgdata/geneeq/geneeqn-emb1/NA1980.DAT.\*  
2: /SID2/gcgdata/geneeq/geneeqn-emb1/NA1981.DAT.\*  
3: /SID2/gcgdata/geneeq/geneeqn-emb1/NA1982.DAT.\*  
4: /SID2/gcgdata/geneeq/geneeqn-emb1/NA1983.DAT.\*  
5: /SID2/gcgdata/geneeq/geneeqn-emb1/NA1984.DAT.\*  
6: /SID2/gcgdata/geneeq/geneeqn-emb1/NA1985.DAT.\*  
7: /SID2/gcgdata/geneeq/geneeqn-emb1/NA1986.DAT.\*  
8: /SID2/gcgdata/geneeq/geneeqn-emb1/NA1987.DAT.\*  
9: /SID2/gcgdata/geneeq/geneeqn-emb1/NA1988.DAT.\*  
10: /SID2/gcgdata/geneeq/geneeqn-emb1/NA1989.DAT.\*  
11: /SID2/gcgdata/geneeq/geneeqn-emb1/NA1990.DAT.\*  
12: /SID2/gcgdata/geneeq/geneeqn-emb1/NA1991.DAT.\*  
13: /SID2/gcgdata/geneeq/geneeqn-emb1/NA1992.DAT.\*  
14: /SID2/gcgdata/geneeq/geneeqn-emb1/NA1993.DAT.\*  
15: /SID2/gcgdata/geneeq/geneeqn-emb1/NA1994.DAT.\*  
16: /SID2/gcgdata/geneeq/geneeqn-emb1/NA1995.DAT.\*  
17: /SID2/gcgdata/geneeq/geneeqn-emb1/NA1996.DAT.\*  
18: /SID2/gcgdata/geneeq/geneeqn-emb1/NA1997.DAT.\*  
19: /SID2/gcgdata/geneeq/geneeqn-emb1/NA1998.DAT.\*  
20: /SID2/gcgdata/geneeq/geneeqn-emb1/NA1999.DAT.\*  
21: /SID2/gcgdata/geneeq/geneeqn-emb1/NA2000.DAT.\*  
22: /SID2/gcgdata/geneeq/geneeqn-emb1/NA2001A.DAT.\*  
23: /SID2/gcgdata/geneeq/geneeqn-emb1/NA2001B.DAT.\*  
24: /SID2/gcgdata/geneeq/geneeqn-emb1/NA2002.DAT.\*

Prod. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	552	100.0	1477	21	AAZ90608
2	105.5	19.1	1713	21	AAZ90608
3	105.5	19.1	1713	21	AAZ90608
4	105	19.0	2197	21	AAZ90608
5	105	19.0	2197	21	AAZ90608
6	105	19.0	2197	21	AAZ90608
7	105	19.0	2197	21	AAZ90608
8	102	18.5	1108	19	AAV44333
9	102	18.5	1108	19	AAV44333
10	102	18.5	1108	20	AAV44333
11	102	18.5	2259	20	AAV44333
12	102	18.5	3757	19	AAV44332
13	102	18.5	3757	19	AAV44332
14	101	18.3	1219	19	AAV44326
15	100	18.1	1154	19	AAV44328
16	100	18.1	1154	19	AAV44328
17	100	18.1	1154	21	AAV44328
18	100	18.1	1154	21	AAV44328
19	100	18.1	5441	21	AAV44328
20	100	18.1	5441	21	AAV44328
21	99	17.9	735	24	AAZ90760
22	99	17.9	735	24	AAZ90760
23	99	17.9	735	24	AAZ90760
24	99	17.9	735	24	AAZ90760
25	99	17.9	735	24	AAZ90760
26	99	17.9	735	24	AAZ90760
27	99	17.9	735	24	AAZ90760
28	99	17.9	735	24	AAZ90760
29	99	17.9	735	24	AAZ90760
30	99	17.9	735	24	AAZ90760
31	99	17.9	735	24	AAZ90760
32	99	17.9	735	24	AAZ90760
33	99	17.9	735	24	AAZ90760
34	99	17.9	735	24	AAZ90760
35	99	17.9	735	24	AAZ90760
36	99	17.9	735	24	AAZ90760
37	99	17.9	735	24	AAZ90760
38	99	17.9	735	24	AAZ90760
39	99	17.9	735	24	AAZ90760
40	99	17.9	735	24	AAZ90760
41	99	17.9	735	24	AAZ90760
42	99	17.9	735	24	AAZ90760
43	99	17.9	735	24	AAZ90760
44	99	17.9	735	24	AAZ90760
45	99	17.9	735	24	AAZ90760

## ALIGNMENTS

RESULT 1	AAZ90608	AAZ90608 standard; DNA, 1477 BP.
ID	AAZ90608	AAZ90608 standard; DNA, 1477 BP.
AC	AAZ90608	AAZ90608 standard; DNA, 1477 BP.
XX	AAZ90608	AAZ90608 standard; DNA, 1477 BP.
DT	19-JUN-2000	(first entry)
XX	19-JUN-2000	(first entry)
DE	S. mansoni	cerarial elastase gene.
XX	S. mansoni	cerarial elastase gene.
KW	Schistosoma parafitei	fusion protein; cercarial elastase; vaccine;
XX	Schistosoma parafitei	fusion protein; cercarial elastase; vaccine;
KW	antibody response; schistosomicide; ds.	antibody response; schistosomicide; ds.
XX	antibody response; schistosomicide; ds.	antibody response; schistosomicide; ds.
OS	Schistosoma mansoni	Schistosoma mansoni
XX	Schistosoma mansoni	Schistosoma mansoni
Key	Location/Qualifiers	Location/Qualifiers
FT	1..1477	1..1477
FT	/*tag= a	/*tag= a

```

FT      /note= "contains introns"
FT      1..168
FT      exon      /tag= b
FT      /number= 1
FT      /note= "encodes the protein fragment shown in AAY67588"
FT      169..309
FT      intron    /tag= c
FT      /number= 1
FT      310..627
FT      exon      /tag= d
FT      /number= 2
FT      /note= "encodes the protein fragment shown in AAY67587"
FT      628..1234
FT      intron    /tag= e
FT      /number= 2
FT      1235..1477
FT      exon      /tag= f
FT      /number= 3
FT      BP992582-A2.
XX      12-APR-2000.
XX      05-OCT-1999; 99BP-0307832.
XX      07-OCT-1998; 98GB-0021821.
XX      (UYWA-) UNITV WALES BANGOR.
XX      Doenhoff M, Sayers J;
XX      MPI; 2000-259136/23.
XX      P-PSDB; AAY67587, AAY67588.
XX      PT New vaccine for treatment of Schistosoma infections contains a
XX      recombinant fusion protein comprising cercarial elastase sequence fused
XX      to bacterial, phage or viral protein -
XX      Claim 2; Fig 1; 26pp; English.
XX      PS The invention provides a vaccine comprising a recombinant fusion protein
XX      capable of eliciting immunity against Schistosoma parasites. The fusion
XX      protein comprises the 27 or 28 kDa cercarial elastase sequence of
XX      S. mansoni or an active fragment, homologue or variant, fused to a
XX      bacterial, phage or viral protein. The vaccine containing the fusion
XX      protein has been found to induce a significantly increased antibody
XX      response against schistosoma infections, compared to the use of S.
XX      CC mansoni cercarial elastase in its native form. The present sequence
XX      represents the S. mansoni cercarial elastase gene.
XX      SQ Sequence 1477 BP; 348 A; 261 C; 402 G; 465 T; 1 other;
XX
XX      Alignment Scores:
XX      Pred. No.: 2.68e-53 Length: 1477
XX      Score: 552.00 Matches: 106
XX      Percent Similarity: 100.00% Conservative: 0
XX      Best Local Similarity: 100.00% Mismatches: 0
XX      Query Match: 100.00% Indels: 0
XX      DB: 21 Gaps: 0
XX
XX      US-10-020-441-2 (1-106) x AAZ90608 (1-1477)
XX
XX      QY 1 ValSerPheLeuThrIleuArgAsnGlyAspGlnGlnGlyIleHisIsgInProSerGly 20
XX      DB 310 GTTTCATTCTTCACACAGAGAAAGGAGCAACCAAGGCAATCCATCCCAACCGTCTGGA 369
XX
XX      QY 21 ValIleValAlaProGlyIleTyrMetProSerGlyMetSerAlaArgIleArgArgProIle 40
XX      DB 370 GTTAAGGTGGACACAGATATCATGCCCTCTTGATGTGCGGACACACAGAGAGCAATC 429
XX
XX      QY 41 AlGlnThrIleuSerGlyIlePheAspIleAlaIleValMetIleuAlaGlnMetValAsnIleu 60
XX      DB 430 GCACAGACACTCAGTGGATTCGATATTCGAATGTAAATGCTGCTCAAAATGTCACACTTA 489

```

```

QY      61 GlnSerGlyIleArgValIleSerIleuProGlnProSerAspIleProProGlyIle 80
DB      490 CAGAGTGAATCAGATGATCATGCTGACACAGCCATCGATATCCGCGACCTGGAAC 549
QY      81 GlyValPheIleValGlyIleTyrGlyIleArgAspAspAspAspArgAspProSerArgIleAsn 100
DB      550 GGTGTTTCATTGTTGGTTATGTGAAGGATGATTAAGACCGTGATCCGTACGTAAGAAT 609
QY      101 GlyGlyIleLeuLysLys 106
DB      610 GGTGAAATATGAGAAA 627
XX
XX      RESULT 2
XX      AAF21084
XX      ID AAF21084 strand: DNA; 17133 BP.
XX      AAF21084;
XX      14-MAR-2001 (first entry)
XX
XX      DE Human low adenosine antisense oligonucleotide related sequence #2651.
XX      XX
XX      KW Low adenosine antisense oligonucleotide; phosphorochiuate; allergy;
XX      human; airway disorder; bronchoconstriction; lung inflammation;
XX      surfactant depletion; respiratory; bronchodilator; antiinflammatory;
XX      immunosuppressive; antiasthmatic; analgesic; hypotensive; cyostatic;
XX      respiratory obstruction; pulmonary obstruction; impeded respiration;
XX      surfactant hypoproduction; pulmonary vasoconstriction; asthma; RDS;
XX      respiratory distress syndrome; pain; cystic fibrosis; allergic rhinitis;
XX      pulmonary hypertension; emphysema; pulmonary transplantation rejection;
XX      chronic obstructive pulmonary disease; pulmonary infection; bronchitis;
XX      cancer; ss.
XX
XX      OS Homo sapiens.
XX      PN WO200062736-A2.
XX      PD 26-OCT-2000.
XX
XX      PF 24-MAR-2000; 2000WO-US08020.
XX      PR 06-APR-1999; 99US-0127958.
XX      PA (UYEC-) UNITV EAST CAROLINA.
XX      PA (NYCE/) NYCE J W.
XX      PI Nyce JW;
XX      DR MPI; 2000-679539/66.
XX
XX      PT Low adenosine (A) content antisense oligonucleotides which do not
XX      trigger adenosine receptors during metabolism, useful e.g. for treating
XX      cancers and respiratory obstructions -
XX
XX      Disclosure; Page 882-887; 1592pp; English.
XX
XX      CC The present invention describes low adenosine (A) content antisense
XX      oligonucleotides and compositions (I) comprising them. In the antisense
XX      oligonucleotides the A is replaced by a 'Universal' or alternative base.
XX      CC (I) can have respiratory, bronchodilator, antiinflammatory, analgesic,
XX      immunosuppressive, antiasthmatic, hypotensive and cyostatic activities.
XX      CC The antisense oligonucleotides and (I) can be used to down-regulate the
XX      expression and or activity of target polypeptides associated with
XX      CC lung/respiratory disorders and malignancies, such as stimulating and
XX      activating peptide factors and transmitters, transcription factors,
XX      CC immunoglobulins and antibodies, antibody receptors, cytokines and
XX      CC chemokines, endogenously produced specific and non-specific enzymes,
XX      CC binding proteins, adhesion molecules and their receptors, cytokine and
XX      CC chemokine receptors, adenosine receptors, bradykinin receptors, central
XX      CC nervous system (CNS) and peripheral nervous and non-nervous system peptide
XX      CC receptors, CNS and peripheral nervous and non-nervous system peptide
XX      CC transmitters, defensins, growth factors, vasoactive peptides and

```

CC receptors, binding proteins and malignancy associated proteins. The  
CC antisense oligonucleotides may be used in this way to treat disorders  
CC including respiratory obstruction (especially pulmonary obstruction  
CC and/or bronchoconstriction) and/or lung inflammation, allergy(ies)  
CC and/or surfactant hypoproduction which are associated with a disease or  
CC condition selected from pulmonary vasoconstriction, inflammation,  
CC allergies, asthma, impeded respiration, respiratory distress syndrome  
CC (RDS), pain, cystic fibrosis (CF), allergic rhinitis (AR), pulmonary  
CC hypertension, emphysema, chronic obstructive pulmonary disease (COPD),  
CC pulmonary transplantation rejection, pulmonary infections, bronchitis,  
CC and/or cancer. AAF1434 to AAR2153 represent human polynucleotide  
CC fragments and antisense oligonucleotides used in the exemplification of  
CC the present invention.

**SQ** Sequence 17133 BP; 2890 A; 5977 C; 5156 G; 3101 T; 9 other;

**Alignment Scores:**

Pred. No.:	0.465	Length:	1713
Score:	105.50	Matches:	32
Percent Similarity:	44.52%	Conservative:	21
Best Local Similarity:	27.12%	Mismatches:	44
Query Match:	19.11%	Indels:	22
DB:	21	Gaps:	3

US-10-020-441-2 (1-106) x AAF21084 (1-17133)

6 LeuArgAsnGlyAspGlnGlnGlyLeuHisGlnProSerGlyValLysValAlaPro 25

Db 11314 CTGGCTCCCGGGTGTCTCTGGGGCTGCCAAGGCCCTGAGTGGATCTCCGCTGCCA 11373

26 Gly-----TyrMetProSerCysMetSerAlaArgGlnArg----- 38

Db 11374 GGGACGTCAGGATCTGGCCGCCCTCAGG-GTGCAACTGCGGAGCAGCACCTCTACTAC 11432

39 -----ProIlealaGlnThrLeu----- 44

DB 11433 CAGGACGAGTGGTGGCCGTACGAGGATCATCGTGCAACCCACAGTCTACACGCCCAG 11434

45 serglypneaspilaeallevaImetleuaIaGIImetValAsnLeuGIInSerglyIle 64

DB 11493 ATCGAGCCGACATCGCCCTGCTGGAGCTGGAGGAGCCCGGTGAACGTCTCCAGCCACGTC 11552

```

05 ArgvallerLeuProGlnProSerAspLeuProProGlyInrGlyValPheIle 84
      :::::|||||:::|||||

```

DD 11333 CACACGGI CACCCICCCCTCAGAGACCTTCCCCCGGGGAI GCCGTCCTGGGTC 11612

03 valgylygi garsu pavadinavimai garsu prielaidos garsu pavadinimai 102

DD 11013 AC100C1GGGCGH1G1GGACAA1GA1GG1GGG1C1GGGACAG1GGAGG1GGG 11060

AAA34962

XX 2025

XX  
170625

21 20-000-2000 (L115C Enclly)  
XX

Human adenosine receptor related polynucleotide SEQ ID NO:2651.

OS Homo sapiens.

PN WO200009525-A2

PD 24-FEB-2000

PF 03-AUG-1999; 99WO-US17712

PR 03-AUG-1998; 98US-0095212

PA (UYEC-) UNIV EAST CAROLINA.

PI NYce JW,

DR WPI; 2000-205971/18.

PT New antisense oligonucleotides useful for treating e.g. pulmonary  
PT vasoconstriction, inflammation, allergies, asthma, hypertension,  
PT bronchitis, emphysema, respiratory distress syndrome, ischemia or  
PT cancers -

PS Disclosure; Page 810-814; 1343pp; English

The present invention describes a new composition comprising an antisense oligonucleotide (ON) with low adenosine (up to 15%), which targets nucleic acids involved in bronchoconstriction, allergies, and/or inflammation. The ON can have anti-inflammatory, antiallergic, antiasthmatic, cyostatic and analgesic activities. The compositions are useful for the treatment of diseases associated with inflammation, impacted airways, including lung disease and diseases whose secondary effects afflict the lungs of a subject. They can be used for treating e.g. ischemic conditions, pulmonary vasoconstriction, allergies, asthma, impacted respiration, respiratory distress syndrome, pain, cystitis, fibrosis, pulmonary hypertension, emphysema, chronic obstructive pulmonary disease (COPD), and cancers such as leukemias, lymphomas, carcinomas, and cancers which may metastasize to the lungs, including breast and prostate cancer. The reduction of the adenosine content of the ONs reduces side effects. The A-containing ONs break down with the release of deoxyadenosine which activates adenosine receptors causing bronchoconstriction and inflammation. AAA2331 to AAA3332 represent the nucleotide sequences given in the sequence listing from the present invention, which correspond to SEQ ID NO:1 to 2615, and then the last 165 sequences are also called SEQ ID NO:1 to 185, but the sequences differ from the previously named sequences. SEQ ID NO:11 to 1680 (AAA33323 to AAA33392) are specifically claimed ONs from the present invention. N.B. Sequences given in the disclosure of the present invention do not match up with their corresponding SEQ ID NO: sequences given in the sequence listing.

SQ Sequence 17133 BP; 2890 A; 5977 C; 5156 G; 3101 T; 9 other;

**Alignment Scores:**

Pred. No.:	0.465	Length:	17133
Score:	105.50	Matches:	32
Percent Similarity:	44.92%	Conservative:	21
Best Local Similarity:	27.12%	Mismatches:	44
Query Match:	19.11%	Indels:	22
DB:	21	Gaps:	3

US-10-020-441-2 (1-106) X AAA34962 (1-17133)

6 LeuArgAsnGlyAspGlnGlnGlyIleHisHisGlnProSerGlyValLysValAlaPro 25

Db 11314 CTGGCTCCCGGCTGCTCCTGGGGCTGCCAGGGCCCTGAGTGGGATCCTCCGCTGCCA 11373

26 Gly-----TyrMetProSerCysMetSerAlaArgGlnArg----- 38

Db 11374 GGGACGTCAGGATCTGGCCGCCCTCAGG-GTGCAACTGGGGAGCAGCACCCTACTAC 11432

QY 39 -----ProIlealagInThrLeu----- 44

Db 11433 CAGGACCAGCTGCTGCCGGTCAGCAGGATCATCTGTCACCCACAGTTCTACACCGCCAG 114

45 SerGlyPheaspIleAlaIleValmetLeuAlaGlnmetValAsnLeuGlnSerGlyIle 64

DB 11493 ATCGAGCGACATCGCCCTGCTGGAGCTGGAGGAGCCGGTGAACGCTCTCCAGCCACGTC 11552

```
Oy 65 ArgValIleSerLeuProGlnProSerAspIleProProGlyThrGlyValPheIle 84
      :::::|||||
Db 11553 CACAGGTCACCTGCCCCCTGCTCAGAGACCTTCCCCCGGAGATGCCGTGCGTC 11612
Oy 85 ValGlyTYrGlyArgAspAspAsnAspArgAspProSerArgIlyAsnGlyIly 102
      :::::|||||
Db 11613 ACTGGCTGGGCGATGTGACATGATGTGGGTCTGGGAGACAGTCAGATGGTGG 11666

RESULT 4
AAF21080
ID AAF21080 standard; DNA; 2197 BP.
XX
XX AAF21080;
AC
XX
XX 14-MAR-2001 (first entry)
DT
XX
XX Human low adenosine antisense oligonucleotide related sequence #2647.
DE
XX low adenosine antisense oligonucleotide; phosphorothioate; allergy;
XX human; airway disorder; bronchoconstriction; lung inflammation;
XX surfactant depletion; respiratory; bronchodilator; antiinflammatory;
XX immunosuppressive; antispasmodic; analgesic; hypotensive; cyostatic;
XX respiratory obstruction; pulmonary obstruction; impeded respiration;
XX surfactant hypoproduction; pulmonary vasoconstriction; asthma; RDS;
XX respiratory distress syndrome; pain; cystic fibrosis; allergic rhinitis;
XX pulmonary hypertension; emphysema; pulmonary transplantation rejection;
XX chronic obstructive pulmonary disease; pulmonary infection; bronchitis;
XX cancer; ss.
XX
XX Homo sapiens.
OS
XX
XX MO200062736-A2.
PN
XX
XX 26-OCT-2000.
PD
XX
XX 24-MAR-2000; 2000WO-US08020.
PF
XX
XX 06-APR-1999; 99US-0127958.
PR
XX
XX (UYEC-) UNIV EAST CAROLINA.
PA
XX (NYCE/) NYCE J W.
PA
XX
XX Nyce JW;
PI
XX
XX WPI; 2000-679539/66.
DR
XX
XX Low adenosine (A) content antisense oligonucleotides which do not
XX trigger adenosine receptors during metabolism, useful e.g. for treating
XX cancers and respiratory obstructions -
XX
XX Disclosure; Page 880-861; 1592pp; English.
XX
XX The present invention describes low adenosine (A) content antisense
XX oligonucleotides and compositions (I) comprising them. In the antisense
XX oligonucleotides the A is replaced by a 'Universal' or alternative base-
XX (I) can have respiratory, bronchodilator, antiinflammatory, analgesic,
XX immunosuppressive, antispasmodic, hypotensive and cyostatic activities.
XX The antisense oligonucleotides and (I) can be used to down-regulate the
XX expression and/or activity of target polypeptides associated with
XX lung/respiratory disorders and malignancies, such as stimulating and
XX activating peptide factors and transmitters, transcription factors,
XX immunoglobulin and antibodies, antibody receptors, cytokines and
XX chemokines, endogenously produced specific and non-specific enzymes,
XX binding proteins, adhesion molecules and their receptors, cytokine and
XX chemokine receptors, adenosine receptors, bradykinin receptors, central
XX nervous system (CNS) and peripheral nervous and non-nervous system
XX receptors, CNS and peripheral nervous and non-nervous system peptide
XX transmitters, defensins, growth factors, vasoactive peptides and
XX receptors, binding proteins and malignancy associated proteins. The
XX antisense oligonucleotides may be used in this way to treat disorders
XX including respiratory obstruction (especially pulmonary obstruction
XX and/or bronchoconstriction) and/or lung inflammation, allergy(ies)
XX and/or surfactant hypoproduction which are associated with a disease or
```

```
CC condition selected from pulmonary vasoconstriction, inflammation,
CC allergies, asthma, impeded respiration, respiratory distress syndrome
CC (RDS), pain, cystic fibrosis (CF), allergic rhinitis (AR), pulmonary
CC hypertension, emphysema, chronic obstructive pulmonary disease (COPD),
CC pulmonary transplantation rejection, pulmonary infections, bronchitis,
CC and/or cancer. AAF18434 to AAF21543 represent human polynucleotide
CC fragments and antisense oligonucleotides used in the exemplification of
CC the present invention.
XX
XX SQ Sequence 2197 BP; 355 A; 789 C; 674 G; 379 T; 0 other;
XX
XX Alignment Scores:
XX Pred. No.: 0.0396 Length: 2197
XX Score: 105.00 Matches: 32
XX Percent Similarity: 44.92% Conservative: 21
XX Best Local Similarity: 27.12% Mismatches: 44
XX Query Match: 19.02% Indels: 22
XX DB: Gaps: 3
XX
XX US-10-020-441-2 (1-106) x AAF21080 (1-2197)
Oy 6 leuArgAsnGlyAspGlnGlnGlyIleHisGlnProSerGlyValIlyValAlaPro 25
      |||
Db 893 CTGGCTCCCGGGTCTCTCTGCGGGGCTGCCAGAGGCCCTGAGTGGATCTCCGCTGCCA 952
Oy 26 Gly-----TyrMetProSerCysMetSerAlaArgGlnArg----- 38
      |||
Db 953 GGGAGCTCAAGATCTGGCGCGGCTTCAGG-GTGCAACTCGGGAGACAGCACTTACTAC 1011
Oy 39 -----ProIleAlaGlnThrLeu----- 44
      |||
Db 1012 CAGGACCACTGCTGCGCTGCGAGATCATGTGACCAACCAAGTCTTACACGCCCGCAG 1071
Oy 45 SerGlyPheAspIleAlaIleValMetLeuAlaGlnMetValAsnLeuGlnSerGlyIle 64
      |||
Db 1072 ATCGAGCGGACATGCGCTGCTGAGCTGAGAGAGGCCGTGAAGCTCTCCAGCCACGTC 1131
Oy 65 ArgValIleSerLeuProGlnProSerAspIleProProGlyThrGlyValPheIle 84
      :::::|||||
Db 1132 CACAGGTCACCTGCCCCCTGCTCAGAGACCTTCCCCCGGAGATGCCGTGCGTCGTC 1191
Oy 85 ValGlyTYrGlyArgAspAspAsnAspArgAspProSerArgIlyAsnGlyIly 102
      :::::|||||
Db 1192 ACTGGCTGGGCGATGTGACATGATGTGGGTCTGGGAGACAGTCAGATGGTGG 1245

RESULT 5
AAA34958
ID AAA34958 standard; DNA; 2197 BP.
XX
XX AAA34958;
AC
XX
XX 28-JUL-2000 (first entry)
DT
XX
XX Human adenosine receptor related polynucleotide SEQ ID NO:2647.
XX
XX Human: adenosine receptor; low adenosine antisense oligonucleotide;
XX phosphorothioate; impaired respiration; inflammation; allergy;
XX allergic disease; bronchoconstriction; inhibitor; antiinflammatory;
XX antiallergic; antispasmodic; cyostatic; analgesic; impaired airway;
XX lung disease; ischaemic condition; pulmonary vasoconstriction; asthma;
XX respiratory distress syndrome; pain; cystic fibrosis; emphysema;
XX pulmonary hypertension; chronic obstructive pulmonary disease; COPD;
XX cancer; leukaemia; lymphoma; carcinoma; metastasis; ss.
XX
XX Homo sapiens.
OS
XX
XX MO200009525-A2.
PN
XX
XX 24-FEB-2000.
PD
XX
XX 03-AUG-1999; 99WO-US17712.
PF
XX
XX 03-AUG-1998; 98US-0095212.
PR
```

XX (UYEC-) UNIV EAST CAROLINA.  
PA  
XX  
XX  
PI Nyce JW;  
XX  
XX WPI; 2000-205971/18.  
DR  
XX  
XX New antisense oligonucleotides useful for treating e.g. pulmonary  
PT vasoconstriction, inflammation, allergies, asthma, hypertension,  
PT bronchitis, emphysema, respiratory distress syndrome, ischemia or  
PT cancers  
XX  
XX  
XX Disclosure; Page 808; 1343bp; English.  
XX  
XX  
XX The present invention describes a new composition comprising an  
CC antisense oligonucleotide (ON) with low adenosine (up to 15%), which  
CC targets nucleic acids involved in bronchoconstriction, allergies, and/or  
CC inflammation. The ON can have anti-inflammatory, antiallergic,  
CC antiasthmatic, cytoskeletal and analgesic activities. The compositions are  
CC useful for the treatment of diseases associated with inflammation,  
CC impaired airways, including lung disease and diseases whose secondary  
CC effects afflict the lungs of a subject. They can be used for treating  
CC e.g. ischemic conditions, pulmonary vasoconstriction, allergies,  
CC asthma, impaired respiration, respiratory distress syndrome, pain, cystic  
CC fibrosis, pulmonary hypertension, emphysema, chronic obstructive  
CC pulmonary disease (COPD), and cancers such as leukemias, lymphomas,  
CC carcinomas, and cancers which may metastasize to the lungs, including  
CC breast and prostate cancer. The reduction of the adenosine content of  
CC the ONs reduces side effects. The A-containing ONs break down with the  
CC release of deoxyadenosine which activates adenosine receptors causing the  
CC bronchoconstriction and inflammation. AAA32313 to AAA35312 represent the  
CC nucleotide sequences given in the sequence listing from the present  
CC invention, which correspond to SEQ ID NO:1 to 2815, and then the last  
CC 185 sequences are also called SEQ ID NO:1 to 185, but the sequences  
CC differ from the previously named sequences. SEQ ID NO:11 to 1680  
CC (AAA32323 to AAA33992) are specifically claimed ONs from the present  
CC invention. N.B. Sequences given in the disclosure of the present  
CC invention do not match up with their corresponding SEQ ID NO: sequences  
CC given in the sequence listing.  
XX  
XX  
XX Sequence 2197 BP; 355 A; 789 C; 674 G; 379 T; 0 other;  
SQ  
Alignment Scores:  
Pred. No.: 0.0396 Length: 2197  
Score: 105.00 Matches: 32  
Percent Similarity: 44.92% Conservative: 21  
Best Local Similarity: 27.12% Mismatches: 44  
Query Match: 19.02% Indels: 22  
DB: Gaps: 3  
US-10-020-441-2 (1-106) x AAA34958 (1-2197)  
QY 6 leuaTgaBenglyAaspGlnGnglylEHSHSGLnProSerGlyVallyValAlaPro 25  
DB 893 CTGGCTCCCGGGTCTCTGGGGGCTGCCAGGGCCCTTAGTGGGATCTCCGCTGCCA.952  
QY 26 GJY-----TyrMetProSerCyMetSerAlaArgGlnrArg----- 38  
DB 953 GGGAGCTCAAGATCTGGCGCCCTTAGG-GTGCACACTGGGAGAGACACTTACTAC 1011  
QY 39 -----ProtleAagInThrLeu----- 44  
DB 1012 CAGGACCACTGCTGCGGTCCGAGAGATCATCTGCACACCAAGTTCTACACCGCCAG 1071  
QY 45 SerGlyPheAspIleAlaIleValMetLeuAlaGlnMetValAsnLeuGlnSerGlyIle 64  
DB 1072 ATCGGAGCGGACATCCCTGCTGGAGCTGGAGAGACCGGTGAACCTCTCCAGCCACGTC 1131  
QY 65 ArgValIleSerLeuProGlnProSerAspIleProProGlyThrGlyValPheIle 84  
DB 1132 CACACGGTACCTGCGCCCTGCTCAGAGACCTTCCCGGGGGATGCCGTCGCTGGGTC 1191  
QY 85 ValGlyTyrGlyArgAspAspAsnAspArgAspProSerArglyAsnGlyGly 102

DB 1192 ACTGGCTGGCGCGATGTGACATGATGTGGTCTGGGAGACATGTGAGTGGG 1245  
RESULT 6  
AAFP21083  
ID AAFP21083 standard; DNA, 2280 BP.  
XX  
XX AAFP21083;  
XX  
XX 14-MAR-2001 (first entry)  
XX  
XX  
XX Human low adenosine antisense oligonucleotide related sequence #2650.  
XX  
XX  
XX Low adenosine antisense oligonucleotide; phosphorothioate; allergy;  
/ human; airway disorder; bronchoconstriction; lung inflammation;  
XX surfactant depletion; respiratory; bronchodilator; antiinflammatory;  
XX immunosuppressive; antiasthmatic; analgesic; hypotensive; cytoskeletal;  
XX respiratory obstruction; pulmonary obstruction; impaired respiration;  
XX surfactant hypoproduction; pulmonary vasoconstriction; asthma; RDS;  
XX respiratory distress syndrome; pain; cystic fibrosis; allergic rhinitis;  
XX pulmonary hypertension; emphysema; pulmonary transplantation rejection;  
XX chronic obstructive pulmonary disease; pulmonary infection; bronchitis;  
XX cancer; ss.  
XX  
XX Homo sapiens.  
OS  
XX  
XX WO20062736-A2.  
PN  
XX 26-OCT-2000.  
PD  
XX  
XX 24-MAR-2000; 2000WO-US08020.  
PF  
XX  
XX 06-APR-1999; 99US-0127958.  
PR  
XX  
XX (UYEC-) UNIV EAST CAROLINA.  
PA  
XX (NYCE/) NYCE J W.  
PI  
XX  
XX Nyce JW;  
XX  
XX WPI; 2000-679539/66.  
DR  
XX  
XX Low adenosine (A) content antisense oligonucleotides which do not  
PT trigger adenosine receptors during metabolism, useful e.g. for treating  
PT cancers and respiratory obstructions -  
XX  
XX  
XX Disclosure; Page 881-882; 1592bp; English.  
XX  
XX The present invention describes low adenosine (A) content antisense  
CC oligonucleotides and compositions (I) comprising them. In the antisense  
CC oligonucleotides the A is replaced by a 'Universal' or alternative base.  
CC (I) can have respiratory, bronchodilator, antiinflammatory, analgesic,  
CC immunosuppressive, antiasthmatic, hypotensive and cytoskeletal activities.  
CC The antisense oligonucleotides and (I) can be used to down-regulate the  
CC expression and or activity of target polypeptides associated with  
CC lung/respiratory disorders and malignancies, such as stimulating and  
CC activating peptide factors and transmitters, transcription factors,  
CC immunoglobulins and antibodies, antibody receptors, cytokines and  
CC chemokines, endogenously produced specific and non-specific enzymes,  
CC binding proteins, adhesion molecules and their receptors, cytokine and  
CC chemokine receptors, adenosine receptors, bradykinin receptors, central  
CC nervous system (CNS) and peripheral nervous and non-nervous system  
CC receptors, CNS and peripheral nervous and non-nervous system peptide  
CC transmitters, defensive, growth factors, vasoactive peptides and  
CC receptors, binding proteins and malignancy associated proteins. The  
CC antisense oligonucleotides may be used in this way to treat disorders  
CC including respiratory obstruction (especially pulmonary obstruction  
CC and/or bronchoconstriction) and/or lung inflammation, allergy(ies)  
CC and/or surfactant hypoproduction which are associated with a disease or  
CC condition selected from pulmonary vasoconstriction, inflammation,  
CC allergies, asthma, impaired respiration, respiratory distress syndrome  
CC (RDS), pain, cystic fibrosis (Cf), allergic rhinitis (AR), pulmonary  
CC hypertension, emphysema, chronic obstructive pulmonary disease (COPD),  
CC pulmonary transplantation rejection, pulmonary infections, bronchitis,



ID AAV44333 standard; DNA; 1108 BP.  
 XX AAV44333;  
 AC  
 XX  
 XX 24-NOV-1998 (first entry)  
 DT  
 XX  
 DE Murine mMCP-6 zymogen nucleic acid sequence.  
 XX  
 XX MCP-6 zymogen; mouse; mast cell protease 6; mMCP-7; tryptase-7;  
 KM blood clot; anticoagulant; myocardial infarction; reocclusion;  
 KM thromboembolism; cerebral embolism; thrombosis; therapy; ss.  
 XX  
 OS Mus sp.  
 XX  
 XX Key Location/Qualifiers  
 FH CDS 34..964  
 FT /\*tag= a  
 FT  
 XX W09824886-A1.  
 XX  
 XX 11-JUN-1998.  
 PD  
 XX 25-NOV-1997; 97WO-US21620.  
 PF  
 XX 04-DEC-1996; 96US-0032354.  
 PR  
 XX (BGHM ) BRIGHAM & WOMENS HOSPITAL.  
 PA  
 XX  
 PI Stevens RL;  
 XX  
 XX WPI; 1998-33308/29.  
 DR P-PSDB; AAM64242.  
 XX  
 XX New compositions containing tryptase-7, e.g. mouse mast cell  
 PT protease-7 - are used to treat clot formation in e.g. myocardial  
 PT infarction, reocclusion following angioplasty or pulmonary  
 PT thromboembolism  
 XX  
 XX Disclosure; Page 70; 92pp; English.  
 XX  
 XX This nucleotide sequence includes a coding region for mouse mast  
 CC cell protease 6 (mMCP-6, see AAM64242). The invention provides:  
 CC compositions comprising an isolated tryptase-7 (such as mMCP-7, see  
 CC AAM64233); a method for treating a blood clot by administering a  
 CC nucleic acid molecule that codes for a tryptase-7, or an expression  
 CC product; a nucleic acid encoding a serine protease (SP); and a  
 CC method of producing a mature SP. Tryptase-7 polypeptides can be  
 CC used for treating disorders mediated by undesirable thrombus clot  
 CC formation such as myocardial infarction and reocclusion following  
 CC angioplasty of blood clots associated with pulmonary  
 CC thromboembolism, deep vein thrombosis, cerebral embolism, renal  
 CC vein and peripheral arterial thrombosis.  
 CC  
 XX  
 XX Sequence 1108 BP; 221 A; 333 C; 297 G; 257 T; 0 other;  
 SQ  
 Alignment Scores:  
 Pred. No.: 0.0367 Length: 1108  
 Score: 102.00 Matches: 21  
 Percent Similarity: 62.50% Conservative: 9  
 Best Local Similarity: 43.75% Mismatches: 18  
 Query Match: 18.48% Indels: 0  
 DB: 19 Gaps: 0  
 US-10-020-441-2 (1-106) x AAV44333 (1-1108)  
 QY 46 G1YPheAp11eAla1leValMe1leuAlaG1Me1ValAsn1leuGlnSerG1Y1leArg 65  
 DB 391 GGGGAGACGTTGCCCTGCTGAGCTTGAAGTCCCTGTAATGTCACACCATATCCAC 450  
 QY 66 Val11eSer1eupProGlnProSerAp11eP1eProProG1Y1ThrG1ValPhe1leVal 85  
 DB 451 CCCATATCCCTGCTCCCTGCTGAGACCTTCCCTGCGACATCTGCTGGTGACA 510

QY 86 G1TYrG1YArgAspAspAsnAsp 93  
 DB 511 GGCTGGGGGACATGTGATATGAC 534  
 RESULT 9  
 ID AAV42709 standard; cDNA; 1108 BP.  
 XX  
 XX AAV42709;  
 AC  
 XX 27-OCT-1998 (first entry)  
 DT  
 XX  
 DE Mouse mast cell protease (mMCP-6) zymogen encoding cDNA.  
 XX  
 XX Mast cell protease; MCP; mouse; inhibitor; peptide substrate; asthma;  
 KM tryptase-6 protein; inflammatory disorder; allergic rhinitis; urticaria;  
 KM antioedema; eczematous dermatitis; atopic dermatitis; anaphylaxis;  
 KM hyperproliferative skin disease; peptic ulcer; hypersensitivity;  
 KM inflammatory skin condition; ss.  
 XX  
 XX Mus sp.  
 OS  
 XX  
 XX Key Location/Qualifiers  
 FH CDS 34..864  
 FT /\*tag= a  
 FT /product= "mMCP-6 zymogen"  
 FT  
 XX W09833812-A1.  
 XX  
 XX 06-AUG-1998.  
 PD  
 XX 30-JAN-1998; 98WO-US01865.  
 PF  
 XX 05-FEB-1997; 97US-0037090.  
 PR  
 XX (BGHM ) BRIGHAM & WOMENS HOSPITAL.  
 PA  
 XX  
 PI Huang C, Stevens RL;  
 XX  
 XX WPI; 1998-437390/37.  
 DR P-PSDB; AAM63172.  
 XX  
 XX Tryptase-6 complex inhibitory peptides - used to treat mast  
 PT cell-mediated inflammatory disorders e.g. asthma  
 PT  
 XX  
 XX Disclosure; Page 43; 69pp; English.  
 XX  
 XX This cDNA encodes the mouse mast cell protease (mMCP-6) zymogen. The  
 CC invention provides sequences shown in AAM63160 to AAM63169 that are  
 CC inhibitors of mMCP-6. These peptides which are tryptase-6 complex  
 CC inhibitors, can be used for treating a mast cell-mediated inflammatory  
 CC disorder. The inhibitors can be used to treat inflammatory disorders  
 CC including asthma, allergic rhinitis, urticaria and antioedema,  
 CC eczematous dermatitis (atopic dermatitis), anaphylaxis,  
 CC hyperproliferative skin disease, peptic ulcers, inflammatory bowel  
 CC disorder, hypersensitivity and inflammatory skin conditions.  
 CC  
 XX  
 XX Sequence 1108 BP; 221 A; 333 C; 297 G; 257 T; 0 other;  
 SQ  
 Alignment Scores:  
 Pred. No.: 0.0367 Length: 1108  
 Score: 102.00 Matches: 21  
 Percent Similarity: 62.50% Conservative: 9  
 Best Local Similarity: 43.75% Mismatches: 18  
 Query Match: 18.48% Indels: 0  
 DB: 19 Gaps: 0  
 US-10-020-441-2 (1-106) x AAV42709 (1-1108)  
 QY 46 G1YPheAp11eAla1leValMe1leuAlaG1Me1ValAsn1leuGlnSerG1Y1leArg 65  
 DB 391 GGGGAGACGTTGCCCTGCTGAGCTTGAAGTCCCTGTAATGTCACACCATATCCAC 450

Oy	66	ValleSeriLeauProGInPrOSeRAspLleProPProGlyThnGLYalPhelIeVal	85
Db	451	CCCAATATCCCTGGCCCCCTCGCAGACTTTCCTCCCTGGACATCGCTGCTGGGTGACA	510
Oy	86	GlyTYrGIYAArgAspaSPasnaSp	93
Db	511	GGCTGGGGCGACATTGATTAATGAC	534
RESULT 10			
ID	AAV33911	standard; cDNA; 2218 BP.	
XX	AAV33911;		
AC	AAV33911;		
XX	12-FEB-1999	(first entry)	
DE	Nucleotide sequence of the trptase-L gene clone NVB352/3.		
KW	Tumour rejection antigen precursor; TRAP; TRH; leukaemia; screening;		
KM	Lymphoma; cancer; HLA; human lymphocyte antigen; vaccine;		
KW	trypase-L; ss.		
OS	Homo sapiens.		
XX			
FH	Key	Location/Qualifiers	
FT	CDS	8..580	
FT		/*tag= a	
PN	M09849299-A1.		
PD	05-NOV-1998.		
PF	22-APR-1998;	98WO-US07784.	
PR	25-APR-1997;	97US-084598.	
XX			
PA	(LUDW-) LUDWIG INST CANCER RES.		
P1	Boon-Falleur T,	Coulie PG, De Smet C, Lucas S, Van Baren N;	
DR	WPI; 1999-009425/01.		
DR	P-PSDB; AAM68540.		
PT	New diagnosis of leukaemia - by detecting genes for tumour antigen		
PT	rejection precursors or corresponding proteins		
PS	Claim 1; Pages 59-61; 88pp; English.		
CC	The present sequence represents the nucleotide sequence of the		
CC	trypase-L gene clone NVB352/3. The protein is a tumour rejection		
CC	antigen precursor (TRAP). The specification describes the treatment of		
CC	disorders which characterised by expression of a leukaemia-associated		
CC	nucleic acid such as TRH. The products are used for in vivo or in vitro		
CC	screening for leukaemia, lymphoma or other cancers by usual		
CC	hybridisation/amplification or immunoassay methods. TRAPs, when		
CC	processed to antigens or complexed with HLA (human lymphocyte antigen)		
CC	molecules, or nucleic acid encoding them, are useful in vaccines for		
CC	treating leukaemia.		
SQ	Sequence 2218 BP; 392 A; 793 C; 659 G; 374 T; 0 other;		
Alignment Scores:			
Pred. No.:	0.0881	Length:	2218
Score:	102.00	Matches:	35
Percent Similarity:	48.57%	Conservative:	16
Best Local Similarity:	33.33%	Mismatches:	35
Query Match:	18.48%	Indels:	19
DB:	20	Gaps:	4
US-10-020-441-2 (1-106) x AAV33911 (1-2218)			
Oy	9	GlyAapGInGLYlYleHisGLInProSGerGLY-----Vallys	22
		::::	

Dd		223	GAGCACGTGCGTGGAAACCCGCATCAAGAATTCTGCCGCCCTCAGGGTGCAACTGCGGGA	282
Oy		23	VAlLAProGIYrMeIProSerCyMeSeSerAIArgINArGaAgProlIEalGlN	42
			:::::	
Dd		283	GCACACACCT---CTACTACCAGGACCAAGCTGTCGCGCATCGAGGATCATCTGGTACCCC	339
Oy		43	ThIReu-----SergIIYPheApIIeAlaIIeValIMethLeuaIagImetVa	58
			:::::	
Dd		340	AoAGTTCTACTATCATCCAGACCGGGGCGGACATCGCCCTGTGTGAGCTGGAGAGCCCCGT	399
Oy		58	IAsnLeugInserGIlyIearValIIeserLeuProGlnProSeraspIIlePropIoPr	78
			:::::	
Dd		400	GAACATCTCCAGCCACATCCACAGGTACAGCTGCCCTCCCTCGGAGAACCTTCCCCCC	459
Oy		78	OGIIThrGLyValpheIIeValGIYrGYLYArgASpaSPasnasPaargaePProSerAr	98
			:::::	
Dd		460	GGGATGCGCGTGGCTGGTCACTGGCTGGGCGCAGCTGGACAAT-----	502
Oy		98	gLyEsAnGlygly 102	
			:::::	
Dd		503	----AAITGTGGG 511	
RESULT 11				
AAV33910				
ID	AAV33910	standard; cDNA; 2259 BP.		
XX				
AC	AAV33910;			
Dt	12-FEB-1999	(first entry)		
XX				
Nucleotide sequence of the trptase-L gene clone NVB352/L.				
DE				
KW	Tumour rejection antigen precursor; TRAP; TRH; leukaemia; screening;			
Lymphoma; cancer; HLA; human lymphocyte antigen; vaccine;				
KM	tryptase-L; ss.			
XX				
OS	Homo sapiens.			
FH	Key	Location/Qualifiers		
FT	CDS	8..580		
FT		/tag= a		
FN	M09849299-A1.			
XX				
PD	05-NOV-1998.			
Pf	22-APR-1998;	98WO-US07784.		
PR	25-APR-1997;	97US-0845998.		
PA	(LUDW-) LUDWIG INST CANCER RES.			
PI	Boon-Falleur T, Coulie PG, De Smet C, Lucas S, Van Baren N;			
DR	WPI; 1999-009425/Ol.			
DR	P-PsDB; AAW68539.			
New diagnosis of leukaemia - by detecting genes for tumour antigen				
rejection precursors or corresponding proteins				
Claim 1; Pages 56-58; 88pp; English.				
The present sequence represents the nucleotide sequence of the				
trypsinase-L gene clone NVB352/L. The protein is a tumour rejection				
antigen precursor (TRAP). The specification describes the treatment of				
diseases which characterised by expression of a leukaemia-associated				
nucleic acid such as TRH. The products are used for in vivo or in vitro				
screening for leukaemia, lymphoma or other cancers by usual				
hybridisation/amplification or immunoassay methods. TRAPS, when				
processed to antigens or complexed with HLA (human lymphocyte antigen)				
molecules, or nucleic acid encoding them, are useful in vaccines for				
treating leukaemia.				



**SQ** Sequence 2259 BP; 391 A; 835 C; 642 G; 391 T; 0 other;

**Alignment Scores:**

Pred. No.:	0.0902	Length:	225
Score:	102.00	Matches:	35
Percent Similarity:	48.57%	Conservative:	16
Best Local Similarity:	33.33%	Mismatches:	35
Query Match:	18.48%	Indels:	19
DB:	20	Gaps:	4

US-10-020-441-2 (1-106) X AAV33910 (1-2259)

[illegible]

RESULT 12  
AAV44332  
ID AAV44332 standard; DNA; 3757 BP

AC	AAV44332;
XX	
DT	24-NOV-1998 (first entry)
XX	
DE	Murine mMCP-6 zymogen gene

KW MCP-6 zymogen; mouse; mast cell protease 6; mMCP-7; trypsinase-7  
KW blood clot; anticoagulant; myocardial infarction; reocclusion;  
KW thromboembolism; cerebral embolism; thrombosis; therapy; ss.

OS	Mus sp.
XX	
FH	Key
FT	CDS
FT	
FT	
FT	exon
FT	
FT	
FT	intron
FT	
FT	
FT	exon
FT	
FT	
FT	intron
FT	
FT	
FT	exon
FT	
FT	
FT	intron
FT	

FT	/number= 3
FT	2697..286C
FT	/tag= h
FT	/number= 4
FT	2861..2933
FT	/tag= i
FT	/number= 4
FT	2934..3098
FT	/tag= j
FT	/number= 5

PN WO9824886-A1.

PD 11-JUN-1998

PF 25-NOV-1997; 97WO-US21620.

PR 04-DEC-1996; 96US-0032354.

PA (BGHM ) BRIGHAM & WOMENS HOSPITAL.

PI Stevens RL;

DR WPI; 1998-333308/29.

2000

PT New coosultions containing tryptase-7, eg. mouse mast cell  
PT protease-7 - are used to treat clot formation in e.g. myocardial  
PT infarction, reocclusion following angioplasty or pulmonary  
PT thromboembolism  
XX  
PS Disclosure, Page 68-69, 92pp; English.

This nucleotide sequence includes a coding region for mouse mast cell protease 6 (mMCP-6, see AAM6424). The invention provides compositions comprising an isolated tryptase-7 (such as mMCP-7, see AAM6423), a method for treating a blood clot by administering a nucleic acid molecule that codes for a tryptase-7, or an expression product; a nucleic acid encoding a serine protease (SP); and a method of producing a mature SP. Tryptase-7 polypeptides can be used for treating disorders mediated by undesirable thrombus clot formation such as myocardial infarction and reocclusion following angioplasty of blood clots associated with pulmonary thromboembolism, deep vein thrombosis, cerebral embolism, renal vein and peripheral arterial thrombosis.

**SD** Sequence 3757 BP; 802 A; 1014 C; 1003 G; 938 T; 0 other;

Alignment Scores:	
Pred. No.:	0.171
Score:	102.00
Percent Similarity:	62.50%
Best Local Similarity:	43.75%
Query Match:	18.48%
BB:	19
length:	375
Matches:	21
Comesative:	9
Mismatches:	18
Indels:	0
Gaps:	0

US-10-020-441-2 (1-106) X AAV44332 (1-3757)

Qy 46 GlyPheaspIleAlaIleValMetLeuAlaGlnMetValAsnLeuGlnSerGlyIleArg 655  
||| |||::|||::: ||| |||::: |||  
Db 2419 GGGGACAGACGTTCCTGCTGGAGCTTGAAGTCCCTGTGAATCTCCACCCATATCCAC 2449

Qy	66	Vall	eser	leu	Pro	Gln	Pro	Ser	Asp	Ile	Pro	Pro	Gly	Thr	Gly	Val	Phe	Ile	Val	85
Db	2479	CCCATAT	CTCC	TGCCCC	CTGC	CGG	AGAC	CTT	CCCCC	CTGG	GAC	ATC	TG	CTG	GGG	TG	AC			2538

QY 86 GLYTYRGlyargAspAspAsnasp 93

Db 2539 GGCTGGGGGCACATTGATAATGAC 2562

RESULT 13  
AAV42708  
ID AAV42708 standard; cDNA; 3757 BP



QY 63 GYILEARVAILLESERLEUPROGINPROSERASPILSEPROBROGLYTHRGYVAL 82  
 DB 588 CATGTCACCCGCTCTCCCTGCTCTGCTGAGACACCTTCCCTCAGGAACATTGTGTC 647  
 QY 83 PHEILEVALIGLYTIRGLYARGASAPASANAAP 93  
 DB 648 TGGGTGACAGCTGGGGAACATCGACAATGAT 680

## RESULT 15

AAV44328 standard; DNA; 1154 BP.

AAV44328;

24-NOV-1998 (first entry)

Human mast cell tryptase alpha nucleic acid sequence.

Mast cell tryptase alpha; human; MCP-7; mast cell protease 7;

blood clot; anticoagulant; myocardial infarction; reocclusion;

thromboembolism; cerebral embolism; thrombosis; therapy; ss.

Homo sapiens.

Key Location/Qualifiers

FT CDS 18..845

FT /\*tag= a

W09824886-A1.

11-JUN-1998.

25-NOV-1997; 97MO-US21620.

04-DEC-1996; 96US-0032354.

(BGHM ) BRIGHAM & WOMENS HOSPITAL.

Stevens RL;

WPI; 1998-33308/29.

P-PSDB; AAM64237.

New compositions containing tryptase-7, e.g. mouse mast cell

protease-7 - are used to treat clot formation in e.g. myocardial

infarction, reocclusion following angioplasty or pulmonary

thromboembolism

Disclosure; Page 63; 92pp; English.

This nucleotide sequence includes a coding region for human

mast cell tryptase alpha (see AAM64237). The invention provides:

compositions comprising an isolated tryptase-7 that may include

chimeric proteins that contain (a) a human tryptase for all but

the active site region and (b) the substrate-binding pocket of

mouse tryptase-7 or its homologues (see AAM64233-36); a method for

treating a blood clot by administering a nucleic acid molecule that

codes for a tryptase-7, or an expression product, to decrease

fibinogen activity; a nucleic acid encoding a serine protease

(SP); and a method of producing a mature SP by expressing the

inactive zymogen in a host cell, and cleaving the enterokinase

susceptibility domain. The tryptase-7 polypeptides can be used to

treat disorders mediated by undesirable thrombus clot formation

such as myocardial infarction and reocclusion following angioplasty

of blood clots associated with pulmonary thromboembolism, deep vein

thrombosis, cerebral embolism, renal vein and peripheral arterial

thrombosis. They are also useful for all surgical procedures that

require decreased blood clots.

Sequence 1154 BP; 203 A; 411 C; 330 G; 210 T; 0 other;

Alignment Scores:

Pred. No.: 0.0652 Length: 1154  
 Score: 100.00 Matches: 28  
 Percent Similarity: 50.00% Conservative: 16  
 Best Local Similarity: 31.82% Mismatches: 39  
 Query Match: 18.12% Indels: 5  
 DB: 19 Gaps: 1

US-10-020-441-2 (1-106) x AAV44328 (1-1154)

QY 11 GINGINGLYILEHIEHIGLINPROSERGLYVALAIPROGLYTHRMEPROSER 30

DB 254 CGTCAGGATTCGGCCACCTCCAGGGTGCACCTGGGAGACACCTTACTACAGCA 313

QY 31 CYMETSERALAARGGLNARGARPROILEALAGINTHLEU-----SERGL 46

DB 314 CCAGCTGCTGCCAGTCACAGCATCATGTGACCCACAGTTCTCATCATCCACACTGG 373

QY 46 YPHEAPILALAILVALMETLEUALGLMEVALANLEUGINSERGLYLEARGVA 66

DB 374 ACCGATATTCGGCTGCTGAGCTGAGAGAGCCGCTGAACATCTCCACCGCTCACAC 433

QY 66 IILESERLEUPROGINPROSERASPILSEPROBROGLYTHRGYVALPHEILEVAL 86

DB 434 GGTGATGCTGCCCTGCTGAGAGACCTTCCCTCGGAGATGCCGTGCTGCTGCTG 493

QY 86 YTRGLYARGASAPASANAAP 93

DB 494 CTGGGCGCATGTCGACAATGAT 515

## RESULT 16

AAV42710 standard; cDNA; 1154 BP.

AAV42710;

27-OCT-1998 (first entry)

Human mast cell tryptase alpha encoding cDNA.

Mast cell protease; MCP; mouse; inhibitor; peptide substrate; asthma;

tryptase-6 protein; inflammatory disorder; allergic rhinitis; urticaria;

antioedema; eczematous dermatitis; atopic dermatitis; anaphylaxis;

hyperproliferative skin disease; peptic ulcer; hypersensitivity;

inflammatory skin condition; human; mast cell tryptase alpha; ss.

Homo sapiens.

Key Location/Qualifiers

FT CDS 18..845

FT /\*tag= a

FT /product= "Human mast cell tryptase alpha"

W0983812-A1.

06-AUG-1998.

30-JAN-1998; 98MO-US01865.

05-FEB-1997; 97US-0037090.

(BGHM ) BRIGHAM & WOMENS HOSPITAL.

Huang C, Stevens RL;

WPI; 1998-437390/37.

P-PSDB; AAM63173.

Tryptase-6 complex inhibitory peptides - used to treat mast

cell-mediated inflammatory disorders e.g. asthma

Disclosure; Pages 44-45; 69pp; English.

This cDNA encodes the human mast cell tryptase alpha which is a



Db 494 CTGGGCGATGTGGACATGAT 515  
 RESULT 18  
 AAA34959  
 ID AAA34959 standard; DNA; 1154 BP.  
 XX  
 AC AAA34959;  
 XX  
 DT 28-JUL-2000 (first entry)  
 XX  
 DE Human adenosine receptor related polynucleotide SEQ ID NO:2648.  
 XX  
 KW Human; adenosine receptor; low adenosine antisense oligonucleotide;  
 KW phosphorothioate; impaired respiration; inflammation; allergy;  
 KW allergic disease; bronchoconstriction; inhibitor; antiinflammatory;  
 KW antiallergic; antiasthmatic; cyostatic; analgesic; impaired airway;  
 KW lung disease; ischaemic condition; pulmonary vasoconstriction; asthma;  
 KW respiratory distress syndrome; pain; cystic fibrosis; emphysema;  
 KW pulmonary hypertension; chronic obstructive pulmonary disease; COPD;  
 KW cancer; leukaemia; lymphoma; carcinoma; metastasis; ss.  
 XX  
 OS Homo sapiens.  
 XX  
 FN WO200009525-A2.  
 XX  
 PD 24-FEB-2000.  
 XX  
 PF 03-AUG-1999; 99WO-US17712.  
 XX  
 PR 03-AUG-1999; 98US-0095212.  
 XX  
 PA (UYEC-) UNIV EAST CAROLINA.  
 XX  
 PI Nyce JW;  
 XX  
 DR WPI; 2000-205971/18.  
 XX  
 PT New antisense oligonucleotides useful for treating e.g. pulmonary  
 PT vasoconstriction, inflammation, allergies, asthma, hypertension,  
 PT bronchitis, emphysema, respiratory distress syndrome, ischemia or  
 PT cancer -  
 XX  
 XX Disclosure; Page 808-809; 1343pp; English.  
 XX  
 CC The present invention describes a new composition comprising an  
 CC antisense oligonucleotide (ON) with low adenosine (up to 15%), which  
 CC targets nucleic acids involved in bronchoconstriction, allergies, and/or  
 CC inflammation. The ON can have antiinflammatory, antiallergic,  
 CC antiasthmatic, cyostatic and analgesic activities. The compositions are  
 CC useful for the treatment of diseases associated with inflammation,  
 CC impaired airways, including lung disease and diseases whose secondary  
 CC effects afflict the lungs of a subject. They can be used for treating  
 CC e.g. ischaemic conditions, pulmonary vasoconstriction, allergies,  
 CC asthma, impeded respiration, respiratory distress syndrome, pain, cystic  
 CC fibrosis, pulmonary hypertension, emphysema, chronic obstructive  
 CC pulmonary disease (COPD), and cancers such as leukemias, lymphomas,  
 CC carcinomas, and cancers which may metastasize to the lungs, including  
 CC breast and prostate cancer. The reduction of the adenosine content of  
 CC the ONs reduces side effects. The A-containing ONs break down with the  
 CC release of deoxyadenosine which activates adenosine receptors causing  
 CC bronchoconstriction and inflammation. AAA3213 to AAA3512 represent the  
 CC nucleotide sequences given in the sequence listing from the present  
 CC invention, which correspond to SEQ ID NO:1 to 2815, and then the last  
 CC 185 sequences are also called SEQ ID NO:1 to 185, but the sequences  
 CC differ from the previously named sequences. SEQ ID NO:11 to 1680  
 CC (AAA3233 to AAA3392) are specifically claimed ONs from the present  
 CC invention. N.B. Sequences given in the disclosure of the present  
 CC invention do not match up with their corresponding SEQ ID NO: sequences  
 CC given in the sequence listing.  
 XX  
 SQ Sequence 1154 BP; 203 A; 411 C; 330 G; 210 T; 0 other;  
 Alignment Scores:

Pred. No.:	0.0652	Length:	1154
Score:	100.00	Matches:	28
Percent Similarity:	50.00%	Conservative:	16
Best Local Similarity:	31.82%	Mismatches:	39
Query Match:	18.12%	Indels:	5
DB:	21	Gaps:	1

US-10-020-441-2 (1-106) x AAA34959 (1-1154)

QY	11	GINGINGLYITLHIEHIEGLINProSerGlyValIAlaProGlyTyrMetProSer	30
DB	254	CGTCAGGATCTGGCCACCTTCAGGGTGCACTGGGGAGCAGCACTTCTACTACCAAGA	313
QY	31	CysMetSerAlaArgIlnArgRProIleAlaGlnThrLeu-----SerG1	46
DB	314	CCAGCTGCTGCCAGTCAGAGATCATCTGCACCAAGTTCTACATCATCAAGACTGG	373
QY	46	yPheApIleAlaIleValMetLeuAlaGlnMetValenLeuGlnSerGlyIleArgVa	66
DB	374	ACCGGATATCGCCCTGCTGAGAGCTGGAGAGCCCGTGAACATCTCCAGCCGCTCACAC	433
QY	66	IlleSerLeuProGlnProSerAspIleProProGlyThrGlyValPheIleValG1	86
DB	434	GGTCATGCTGCCCTGCTCGAGACCTTCCCGGGGATGCCGTGCTGCTCACTGG	493
QY	86	yTyrGlyArgAspAspAspAsp 93	
DB	494	CTGGGCGATGTGGACATGAT 515	

RESULT 19  
 AAF21074  
 ID AAF21074 standard; DNA; 5441 BP.  
 XX  
 AC AAF21074;  
 XX  
 DT 14-MAR-2001 (first entry)  
 XX  
 DE Human low adenosine antisense oligonucleotide related sequence #2641.  
 XX  
 KW Low adenosine antisense oligonucleotide; phosphorothioate; allergy;  
 KW human; airway disorder; bronchoconstriction; lung inflammation;  
 KW surfactant depletion; respiratory; bronchodilator; antiinflammatory;  
 KW immunosuppressive; antiasthmatic; analgesic; hypotensive; cyostatic;  
 KW respiratory obstruction; pulmonary obstruction; impeded respiration;  
 KW surfactant hypoproduction; pulmonary vasoconstriction; asthma; RDS;  
 KW respiratory distress syndrome; pain; cystic fibrosis; allergic rhinitis;  
 KW pulmonary hypertension; emphysema; pulmonary transplantation rejection;  
 KW chronic obstructive pulmonary disease; pulmonary infection; bronchitis;  
 KW cancer; ss.  
 XX  
 OS Homo sapiens.  
 XX  
 FN WO200062736-A2.  
 XX  
 PD 26-OCT-2000.  
 XX  
 PF 24-MAR-2000; 2000WO-US08020.  
 XX  
 PR 06-APR-1999; 99US-0127958.  
 XX  
 PA (UYEC-) UNIV EAST CAROLINA.  
 PA (NYCE/) NYCE J W.  
 XX  
 PI Nyce JW;  
 XX  
 DR WPI; 2000-679539/66.  
 XX  
 PT Low adenosine (A) content antisense oligonucleotides which do not  
 PT trigger adenosine receptors during metabolism, useful e.g. for treating  
 PT cancers and respiratory obstructions -  
 XX  
 PS Disclosure; Page 877-878; 1592pp; English.  
 XX

The present invention describes low adenosine (A) content antisense oligonucleotides and compositions (I) comprising them. In the antisense oligonucleotides the A is replaced by a 'universal' or alternative base (I) can have respiratory, bronchodilator, antiinflammatory, analgesic, immunosuppressive, antisthmatic, hypotensive and cytostatic activities. The antisense oligonucleotides and (I) can be used to down-regulate the expression and/or activity of target polypeptides associated with lung/respiratory disorders and malignancies, such as stimulating and activating peptide factors and transmitters, transcription factors, immunoglobulins and antibodies, antibody receptors, cytokines and chemokines, endogenously produced specific and non-specific enzymes, binding proteins, adhesion molecules and their receptors, cytokine and chemokine receptors, adenosine receptors, bradykinin receptors, central nervous system (CNS) and peripheral nervous and non-nervous system receptors, CNS and peripheral nervous and non-nervous system peptide transmitters, defensins, growth factors, vasoactive peptides and receptors, binding proteins and malignancy associated proteins. The antisense oligonucleotides may be used in this way to treat disorders including respiratory obstruction (especially pulmonary obstruction and/or bronchoconstriction) and/or lung inflammation, allergy(ies) and/or surfactant hypoproduction which are associated with a disease or condition selected from pulmonary vasoconstriction, inflammation, allergies, asthma, impeded respiration, respiratory distress syndrome (RDS), pain, cystic fibrosis (CF), allergic rhinitis (AR), pulmonary hypertension, emphysema, chronic obstructive pulmonary disease (COPD), pulmonary transplantation rejection, pulmonary infections, bronchitis, and/or cancer. APr1834 to APr2153 represent human polynucleotide fragments and antisense oligonucleotides used in the exemplification of the present invention.

SQ Sequence 5441 BP; 952 A; 1754 C; 1694 G; 1032 T; 9 other;

Alignment Scores:

Pred. No.:	0.463	Length:	5441
Score:	100.00	Matches:	30
Percent Similarity:	46.79%	Conservative:	21
Best Local Similarity:	27.52%	Mismatches:	37
Query Match:	18.12%	Indels:	22
DB:	21	Gaps:	3

US-10-020-441-2 (1-106) X AAF21074 (1-5441)

```

QY      6  LeuArganGlyArgGlnGlnGlyIleHisGlnProSerGlyValIlyValAlaPro 25
Db      2552  CTGGCTCCCGGAGCTGCTCTGGAGGAGCTCCACGGGACCTTAGTGGAGATCTCCGCTGCCA 2611
QY      26  Gly-----TyrMetProSerCysMetSerAlaArgIleArgArg----- 38
Db      2612  GGGAGCTCAAGATCTGGCCACCTTCAGG-GTCCAACTCGGAGACAGACCTTACTAC 2670
QY      39  -----ProIleAlaGlnThrLeu----- 44
Db      2671  CAGGACCAAGCTGCTGCCGGTCAAGAGATATGTCACACCCACAGTTTACTATCATCAG 2730
QY      45  SerGlyPheaspIleAlaIleValMetLeuAlaGlnMetValAsnIeuGlnSerGlyIle 64
Db      2731  ACTGGAGCGGATATTCGCGCTGCTGAGACTGGAGAGCCGTGAACATCTCAGCGCGCTC 2790
QY      65  ArgValIleSerLeuProGlnProSerAspIleProProGlyIThrGlyValPheIle 84
Db      2791  CACACGCTCAGCTGCGCCCTCGCTCGAGAACCTTCCCCCGGGGATGCGCTGCTGGTTC 2850
QY      85  ValGlyIlyTrgIlyArgAspAspAsp 93
Db      2851  ACTGGCTGGGGCGAGTGTGGACAATGAAT 2877

```

```
RESULT 20
AAA34952
ID AAA34952 standard; DNA; 5441 BP
XX
AC AAA34952;
XX
DT 28-JUL-2000 (first entry)
```

Human adenosine receptor related polynucleotide SEQ ID NO:2641.

Human, adenosine receptor; low adenosine antisense oligonucleotide; phosphorothioate; impaired respiration; inflammation; allergy; allergic disease; bronchoconstriction; inhibitor; antiinflammatory; antiallergic; antisthmatic; cyostatic; analgesic; impaired airway; lung disease; ischemic condition; pulmonary vasoconstriction; asthma; respiratory distress syndrome; pain; cystic fibrosis; emphysema; pulmonary hypertension; chronic obstructive pulmonary disease; COPD; Cancer; Leukaemia; Lymphoma; Carcinoma; metastasis; ss.

Homo sapiens.

WO200009525-A2.

24-FEB-2000.

03-AUG-1999; 99WO-US17712.

03-AUG-1998; 98US-0095212.

(UYEC-) UNIV EAST CAROLINA.

Nyce JW;

WPI; 2000-205971/18.

New antisense oligonucleotides useful for treating e.g. pulmonary vasoconstriction, inflammation, allergies, asthma, hypertension, bronchitis, emphysema, respiratory distress syndrome, ischemia or cancers -

Disclosure; Page 805-806; 1343pp; English.

The present invention describes a new composition comprising an antisense oligonucleotide (ON) with low adenosine (up to 15%), which targets nucleic acids involved in bronchoconstriction, allergies, and/or inflammation. The ON can have antiinflammatory, antiallergic, antiasthmatic, cyostatic and analgesic activities. The compositions are useful for the treatment of diseases associated with inflammation, impaired airways, including lung disease and diseases whose secondary effects afflict the lungs of a subject. They can be used for treating e.g. ischemic conditions, pulmonary vasoconstriction, allergies, asthma, impeded respiration, respiratory distress syndrome, pain, cystic fibrosis, pulmonary hypertension, emphysema, chronic obstructive pulmonary disease (COPD), and cancers such as leukemias, lymphomas, carcinomas, and cancers which may metastasize to the lungs, including breast and prostate cancer. The reduction of the adenosine content of the ONs reduces side effects. The A-containing ONs break down with the release of deoxyadenosine which activates adenosine receptors causing bronchoconstriction and inflammation. AAA32313 to AAA3512 represent the nucleotide sequences given in the sequence listing from the present invention, which correspond to SEQ ID NO:1 to 2815, and then the last 185 sequences are also called SEQ ID NO:1 to 185, but the sequences differ from the previously named sequences. SEQ ID NO:11 to 1680 (AAA32323 to AAA33992) are specifically claimed ONs from the present invention. N.B. Sequences given in the disclosure of the present invention do not match up with their corresponding SEQ ID NO: sequences given in the sequence listing.

Sequence 5441 BP; 952 A; 1754 C; 1694 G; 1032 T; 9 other;

SQ Sequence 5441 BP; 952 A, 1754 C, 1694 G, 1032 T; 9 other	
Alignment Scores:	
Pred. No.:	0.463
Score:	100.00
Percent Similarity:	46.79%
Best Local Similarity:	27.52%
Query Match:	18.12%
DB:	21
	3
	5441
Length:	
Matches:	30
Conservative:	21
Mismatches:	37
Indels:	22
Gaps:	3

US-10-020-441-2 (1-106) X AAA34952 (1-5441)

Qy 6 LeuArganGlyAspGlnGlnGlyIleHisGlnProSerGlyValIysValAlaPro 25  
 Db 2552 CTGGCTCCCGGGGTGCTCTCGTGGGGGCTGCCAGAGCCCTGAGTGGGATCTCCGCTGCCA 2611  
 Qy 26 Gly-----TyrMetProSerGlyMetSerAlaArgGlnArgArg----- 38  
 Db 2612 GGGAGCTCAAGATCTGGCCACCTTCAAG-GTGGCAACTCGGGAGACGACCACTTACTAC 2670  
 Qy 39 -----ProIleAlaGlnThrLeu----- 44  
 Db 2671 CAGGACCAAGCTCTCCGCTCAGCAGGATCATCTGCACCCACAGTTTACTATCATCCAG 2730  
 Qy 45 SerGlyPheAspIleAlaIleValMetLeuAlaGlnMetValAsnLeuGlnSerGlyIle 64  
 Db 2731 ACTGGAGCGGATATGCGCTGCTGGAGCTGAGAGCCCGGTAAACATCTCCAGCCGCGTC 2790  
 Qy 65 ArgValIleSerLeuProGlnProSerAspIleProProGlyThrGlyValPheIle 84  
 Db 2791 CACACGGTCATGCTGCCCCCTGCTCGGAGACCTTCCGCCGGGATGCCGTGTGGGTC 2850  
 Qy 85 ValGlyTyrGlyArgAspAspAsp 93  
 Db 2851 ACTGGCTGGCGGATGTGACATGAT 2877

## RESULT 21

AAS20760  
 ID AAS20760 standard; DNA; 735 BP.

XX AAS20760;

DT 09-APR-2002 (first entry)

XX DNA encoding human mature beta-I tryptase.

KW Human; proteolytic tryptase; protease; mature beta-I tryptase;

KM enzyme; gene; ds.

XX Homo sapiens.

OS Location/Qualifiers

FT Key 1..735

FT CDS /\*tag= a

FT /partial

FT /product= "Mature beta-I tryptase"

FT /note= "This sequence lacks both start and stop codons"

XX WO200198470-A2.

XX 27-DEC-2001.

XX 20-JUN-2001; 2001WO-US19681.

XX 21-JUN-2000; 2000US-0598982.

XX (PROM-) PROMEGA CORP.

XX Maffit M, Niles AL, Haak-Frendescho M;

XX WPI; 2002-114578/15.

XX P-PSDB; AAU12005.

XX DNA construct for producing enzymatically-inactive proteolytic

PT tryptase, comprises DNA sequence encoding proteolytic tryptase having

PT an active site mutation -

XX Disclosure; Page 69-70; 126pp; English.

XX The present invention relates to recombinant human proteolytic  
 CC tryptases, active site mutants of these tryptases and the methods for  
 CC producing these. The method involves the production of a DNA expression  
 CC construct comprising a promoter operably linked to a secretion signal  
 CC sequence which is operably linked to a DNA sequence encoding a  
 CC proteolytic tryptase with an active site mutation (the construct drives

CC expression of a mature proteolytic tryptase that lacks enzymatic activity  
 CC due to the active site mutation, in hosts transformed to contain the  
 CC construct). The method is useful for producing enzymatically-active  
 CC beta-II tryptase. The active site mutants of proteolytic tryptase provide  
 CC a tool to investigate the structural and functional properties of the  
 CC protease and its enzymatic activity, and for modelling studies. The  
 CC enzymatically-active, recombinant proteolytic tryptase produced are  
 CC useful as an antigen to generate anti-human tryptase antibodies  
 CC and in drug screening for compounds which act as tryptase inhibitors,  
 CC antagonists, agonists, etc. The present sequence encodes for human  
 CC mature beta-I tryptase.

XX SQ Sequence 735 BP; 141 A; 246 C; 228 G; 120 T; 0 other;

## Alignment Scores:

Score:	0.048	Length:	735
Percent Similarity:	99.00	Matches:	19
Best Local Similarity:	62.50%	Conservative:	11
Query Match:	39.58%	Mismatches:	18
DB:	17.93%	Indels:	0
	24	Gaps:	0

US-10-020-441-2 (1-106) x AAS20760 (1-735)

Qy 46 GlyPheAspIleAlaIleValMetLeuAlaGlnMetValAsnLeuGlnSerGlyIleArg 65  
 Db 265 GGGAGCGGACATGCGCTGCTGGAGCTGAGAGCGCGTGAACGTCTCCAGCAGCTCCAC 324

Qy 66 ValIleSerLeuProGlnProSerAspIleProProGlyThrGlyValPheIleVal 85  
 Db 325 ACGGTACACCTGCCCCCTGCTCAGAGACCTTCCGCCGGGATGCCGTGTGGGTC 384

Qy 86 GlyTyrGlyArgAspAspAsp 93  
 Db 385 GCGTGGCGGATGTGACATGAT 408

## RESULT 22

AAS20787  
 ID AAS20787 standard; DNA; 735 BP.

XX AAS20787;

DT 09-APR-2002 (first entry)

XX DNA encoding human beta-II tryptase active site mutant H44A #4.

DE Human; proteolytic tryptase; protease; recombinant beta-II tryptase;

KW enzyme; mutant; ds.

XX Homo sapiens.

OS Synthetic.

XX WO200198470-A2.

XX 27-DEC-2001.

XX 20-JUN-2001; 2001WO-US19681.

XX 21-JUN-2000; 2000US-0598982.

XX (PROM-) PROMEGA CORP.

XX Maffit M, Niles AL, Haak-Frendescho M;

XX WPI; 2002-114578/15.

XX P-PSDB; AAU12021.

XX DNA construct for producing enzymatically-inactive proteolytic

PT tryptase, comprises DNA sequence encoding proteolytic tryptase having

PT an active site mutation -

XX Example 1c; Page 115-117; 126pp; English.

CC The present invention relates to recombinant human proteolytic  
CC tryptases, active site mutants of these tryptases and the methods for  
CC producing these. The method involves the production of a DNA expression  
CC construct comprising a promoter operably linked to a secretion signal  
CC sequence which is operably linked to a DNA sequence encoding a  
CC proteolytic tryptase with an active site mutation (the construct drives  
CC expression of a mature proteolytic tryptase that lacks enzymatic activity  
CC due to the active site mutation, in hosts transformed to contain the  
CC construct). The method is useful for producing enzymatically-active  
CC beta-II tryptase. The active site mutants of proteolytic tryptase provide  
CC a tool to investigate the structural and functional properties of the  
CC protease and its enzymatic activity, and for modelling studies. The  
CC enzymatically-active, recombinant proteolytic tryptase produced are  
CC useful as an antigen to generate anti-human tryptase antibodies  
CC and in drug screening for compounds which act as tryptase inhibitors,  
CC antagonists, agonists, etc. AAS20775-AAS20790 encode for recombinant  
CC human beta-II tryptase active site mutants.

XX SQ Sequence 735 BP; 139 A; 246 C; 230 G; 120 T; 0 other;

Alignment Scores:  
Pred. No.: 0.048 Length: 735  
Score: 99.00 Matches: 19  
Percent Similarity: 62.50% Conservative: 11  
Best Local Similarity: 39.58% Mismatches: 18  
Query Match: 17.93% Indels: 0  
DB: 24 Gaps: 0

US-10-020-441-2 (1-106) x AAS20787 (1-735)

QY 46 G|YpHeAsp|LeAla|IleValMetLeuAlaGlnMetValAsnLeuGlnSerGlyLeArg 65  
DB 265 GGAGGGGAGCATGCGCCCTGCTGGAGCTGGAGAGCGCGTGAACGTCTCCAGCCACGTCAC 324

QY 66 Val|LeSerLeuProGlnProSerAsp|LeProProGlyThrc|yValPhe|leVal 85  
DB 335 ACGTCAACCTGCCCCCTGCTCAGAGACCTTCCCCCGGGGATGCGCTGGGTCAC 384

QY 86 G|YTYRGLA|ArgAspAspAsp 93  
DB 385 GGCTGGGGCGATGTGACATGAT 408

RESULT 23

AAS20789 ID AAS20789 standard; DNA; 735 BP.

AC AAS20789;

DT 09-APR-2002 (first entry)

DE DNA encoding human beta-II tryptase active site mutant S194A #7.

XX Human; proteolytic tryptase; protease; recombinant beta-II tryptase;  
KW enzyme; mutant; ds.

XX Homo sapiens.  
OS Synthetic.

XX WO200198470-A2.

XX PD 27-DEC-2001.

XX PF 20-JUN-2001; 2001WO-US19681.

XX PR 21-JUN-2000; 2000US-0598982.

XX PA (PROM-) PROMEGA CORP.

XX PI Maffit M, Niles AL, Haak-Frendscho M;

XX DR WPI; 2002-114578/15.

XX P-PSDB; AAU12023.

XX

PT DNA construct for producing enzymatically-inactive proteolytic  
PT tryptase, comprises DNA sequence encoding proteolytic tryptase having  
PT an active site mutation -  
XX  
XX Example 1c; Page 121-122; 126pp; English.

XX The present invention relates to recombinant human proteolytic  
CC tryptases, active site mutants of these tryptases and the methods for  
CC producing these. The method involves the production of a DNA expression  
CC construct comprising a promoter operably linked to a secretion signal  
CC sequence which is operably linked to a DNA sequence encoding a  
CC proteolytic tryptase with an active site mutation (the construct drives  
CC expression of a mature proteolytic tryptase that lacks enzymatic activity  
CC due to the active site mutation, in hosts transformed to contain the  
CC construct). The method is useful for producing enzymatically-active  
CC beta-II tryptase. The active site mutants of proteolytic tryptase provide  
CC a tool to investigate the structural and functional properties of the  
CC protease and its enzymatic activity, and for modelling studies. The  
CC enzymatically-active, recombinant proteolytic tryptase produced are  
CC useful as an antigen to generate anti-human tryptase antibodies  
CC and in drug screening for compounds which act as tryptase inhibitors,  
CC antagonists, agonists, etc. AAS20775-AAS20790 encode for recombinant  
CC human beta-II tryptase active site mutants.

SQ Sequence 735 BP; 142 A; 245 C; 227 G; 121 T; 0 other;

Alignment Scores:  
Pred. No.: 0.048 Length: 735  
Score: 99.00 Matches: 19  
Percent Similarity: 62.50% Conservative: 11  
Best Local Similarity: 39.58% Mismatches: 18  
Query Match: 17.93% Indels: 0  
DB: 24 Gaps: 0

US-10-020-441-2 (1-106) x AAS20789 (1-735)

QY 46 G|YpHeAsp|LeAla|IleValMetLeuAlaGlnMetValAsnLeuGlnSerGlyLeArg 65  
DB 265 GGAGGGGAGCATGCGCCCTGCTGGAGCTGGAGAGCGCGTGAACGTCTCCAGCCACGTCAC 324

QY 66 Val|LeSerLeuProGlnProSerAsp|LeProProGlyThrc|yValPhe|leVal 85  
DB 335 ACGTCAACCTGCCCCCTGCTCAGAGACCTTCCCCCGGGGATGCGCTGGGTCAC 384

QY 86 G|YTYRGLA|ArgAspAspAsp 93  
DB 385 GGCTGGGGCGATGTGACATGAT 408

RESULT 24

AAS20790 ID AAS20790 standard; DNA; 735 BP.

AC AAS20790;

DT 09-APR-2002 (first entry)

DE DNA encoding human beta-II tryptase active site mutant S194A #8.

XX Human; proteolytic tryptase; protease; recombinant beta-II tryptase;  
KW enzyme; mutant; ds.

XX Homo sapiens.  
OS Synthetic.

XX WO200198470-A2.

XX PD 27-DEC-2001.

XX PF 20-JUN-2001; 2001WO-US19681.

XX PR 21-JUN-2000; 2000US-0598982.

XX PA (PROM-) PROMEGA CORP.



```

XX Maffit M, Niles AL, Haak-Frendescho M;
PI WPI; 2002-114578/15.
DR P-PSDB; AAU12024.
XX
PT DNA construct for producing enzymatically-inactive proteolytic
PT trypsinase, comprises DNA sequence encoding proteolytic trypsinase having
PT an active site mutation -
XX
PS Example 1c; Page 124-125; 126pp; English.
XX
CC The present invention relates to recombinant human proteolytic
CC trypsinases, active site mutants of these trypsinases and the methods for
CC producing these. The method involves the production of a DNA expression
CC construct comprising a promoter operably linked to a secretion signal
CC sequence which is operably linked to a DNA sequence encoding a
CC proteolytic trypsinase with an active site mutation (the construct drives
CC expression of a mature proteolytic trypsinase that lacks enzymatic activity
CC due to the active site mutation, in hosts transformed to contain the
CC construct). The method is useful for producing enzymatically-active
CC beta-II trypsinase. The active site mutants of proteolytic trypsinase provide
CC a tool to investigate the structural and functional properties of the
CC trypsinase and its enzymatic activity, and for modelling studies. The
CC enzymatically-active, recombinant proteolytic trypsinase produced are
CC useful as an antigen to generate anti-human trypsinase antibodies
CC and in drug screening for compounds which act as trypsinase inhibitors,
CC antagonists, agonists, etc. AAS20775-AAS20790 encode for recombinant
CC human beta-II trypsinase active site mutants.
XX
SQ Sequence 735 BP; 144 A; 245 C; 227 G; 119 T; 0 other;
XX
Alignment Scores:
Pred. No.: 0 048 Length: 735
Score: 99.00 Matches: 19
Percent Similarity: 62.50% Conservative: 11
Best Local Similarity: 39.58% Mismatches: 18
Query Match: 17.93% Indels: 0
DB: 24 Gaps: 0
US-10-020-441-2 (1-106) x AAS20790 (1-735)
Qy 46 GlyPheAspIleAlaIleValMetLeuValAsnLeuGlnSerGlyIleArg 65
Db 265 GGAGCGGACATCGCCCTCTGAGCTGAGAGAGCCGGTAAAGTCTCCAGCCACATCCAC 324
Qy 66 ValIleSerLeuProGlnProSerAspIleProProGlyThrGlyValPheIleVal 85
Db 325 ACGGTACACCTGCCCCCTGCTCAGAGACCTTCCCCCGGGGATGCGGTGCTGACT 384
Qy 86 GlyTyrGlyArgAspAspAsp 93
Db 385 GGCTGGGCGATGTGACATGAT 408
XX
RESULT 25
AAS20763
ID AAS20763 standard; DNA; 771 BP.
XX
AC AAS20763;
XX
DT 09-APR-2002 (first entry)
XX
DE DNA encoding human beta-I trypsinase.
XX
KW Human; proteolytic trypsinase; protease; beta-I trypsinase;
XX enzyme; gene; ds.
XX
OS Homo sapiens.
XX
FH Key
FT CDS Location/Qualifiers
FT 7..756
FT /*tag=a
FT /partial

```

```

FT /product="Beta-I trypsinase"
FT /note="This sequence lacks a start codon"
XX
XX WO200198470-A2.
XX
XX 27-DEC-2001.
XX
XX 20-JUN-2001; 2001WO-US19681.
XX
XX 21-JUN-2000; 2000US-0598982.
XX
XX (PROM-) PROMEGA CORP.
XX
XX Maffit M, Niles AL, Haak-Frendescho M;
XX WPI; 2002-114578/15.
XX P-PSDB; AAU12006.
XX
XX DNA construct for producing enzymatically-inactive proteolytic
XX trypsinase, comprises DNA sequence encoding proteolytic trypsinase having
XX an active site mutation -
XX
PS Disclosure; Page 72-74; 126pp; English.
XX
CC The present invention relates to recombinant human proteolytic
CC trypsinases, active site mutants of these trypsinases and the methods for
CC producing these. The method involves the production of a DNA expression
CC construct comprising a promoter operably linked to a secretion signal
CC sequence which is operably linked to a DNA sequence encoding a
CC proteolytic trypsinase with an active site mutation (the construct drives
CC expression of a mature proteolytic trypsinase that lacks enzymatic activity
CC due to the active site mutation, in hosts transformed to contain the
CC construct). The method is useful for producing enzymatically-active
CC beta-II trypsinase. The active site mutants of proteolytic trypsinase provide
CC a tool to investigate the structural and functional properties of the
CC trypsinase and its enzymatic activity, and for modelling studies. The
CC enzymatically-active, recombinant proteolytic trypsinase produced are
CC useful as an antigen to generate anti-human trypsinase antibodies
CC and in drug screening for compounds which act as trypsinase inhibitors,
CC antagonists, agonists, etc. The present sequence encodes for human
CC beta-I trypsinase.
XX
SQ Sequence 771 BP; 149 A; 257 C; 241 G; 124 T; 0 other;
XX
Alignment Scores:
Pred. No.: 0 051 Length: 771
Score: 99.00 Matches: 19
Percent Similarity: 62.50% Conservative: 11
Best Local Similarity: 39.58% Mismatches: 18
Query Match: 17.93% Indels: 0
DB: 24 Gaps: 0
US-10-020-441-2 (1-106) x AAS20763 (1-771)
Qy 46 GlyPheAspIleAlaIleValMetLeuValAsnLeuGlnSerGlyIleArg 65
Db 283 GGAGCGGACATCGCCCTCTGAGCTGAGAGAGCCGGTAAAGTCTCCAGCCACATCCAC 342
Qy 66 ValIleSerLeuProGlnProSerAspIleProProGlyThrGlyValPheIleVal 85
Db 343 ACGGTACACCTGCCCCCTGCTCAGAGACCTTCCCCCGGGGATGCGGTGCTGACT 402
Qy 86 GlyTyrGlyArgAspAspAsp 93
Db 403 GGCTGGGCGATGTGACATGAT 426
XX
RESULT 26
AAS20783
ID AAS20783 standard; DNA; 771 BP.
XX
AC AAS20783;
XX
DT 09-APR-2002 (first entry)

```

XX DNA encoding human beta-II tryptase active site mutant H44A #3.  
DE  
XX  
XX Human; proteolytic tryptase; protease; recombinant beta-II tryptase;  
KW enzyme; mutant; ds.  
XX  
XX Homo sapiens.  
OS Synthetic.  
XX  
XX WO200198470-A2.  
XX  
XX  
XX 27-DEC-2001.  
XX  
XX 20-JUN-2001; 2001WO-US19681.  
XX  
XX 21-JUN-2000; 2000US-0598982.  
XX  
XX (PROM-) PROMEGA CORP.  
XX  
XX Maffit M, Niles AL, Haak-Frendescho M;  
XX  
XX MPI; 2002-114578/15.  
XX  
XX P-PSDB; AAU12017.  
XX  
XX  
XX DNA construct for producing enzymatically-inactive proteolytic  
PT tryptase, comprises DNA sequence encoding proteolytic tryptase having  
PT an active site mutation -  
XX  
XX  
XX Claim 7; Page 104-105; 126pp; English.  
XX  
XX The present invention relates to recombinant human proteolytic  
CC tryptases, active site mutants of these tryptases and the methods for  
CC producing these. The method involves the production of a DNA expression  
CC construct comprising a promoter operably linked to a secretion signal  
CC sequence which is operably linked to a DNA sequence encoding a  
CC proteolytic tryptase with an active site mutation (the construct drives  
CC expression of a mature proteolytic tryptase that lacks enzymatic activity  
CC due to the active site mutation, in hosts transformed to contain the  
CC construct). The method is useful for producing enzymatically-active  
CC beta-II tryptase. The active site mutants of proteolytic tryptase provide  
CC a tool to investigate the structural and functional properties of the  
CC protease and its enzymatic activity, and for modelling studies. The  
CC enzymatically-active, recombinant proteolytic tryptase produced are  
CC useful as an antigen to generate anti-human tryptase antibodies  
CC and in drug screening for compounds which act as tryptase inhibitors,  
CC antagonists, agonists, etc. AAS20775-AAS20790 encode for recombinant  
CC human beta-II tryptase active site mutants.  
XX  
XX  
SQ Sequence 771 BP; 147 A; 257 C; 243 G; 124 T; 0 other;  
XX  
XX  
XX Alignment Scores:  
Pred. No.: 0.051 Length: 771  
Score: 99.00 Matches: 19  
Percent Similarity: 62.50% Conservative: 11  
Best Local Similarity: 39.58% Mismatches: 18  
Query Match: 17.93% Indels: 0  
DB: 24 Gaps: 0  
US-10-020-441-2 (1-106) x AAS20783 (1-771)  
QY 46 GYPhneAsp11eAla1eValMeLeuAlaG1MeCValAsnLeuGlnSerGly1IeArg 65  
DB 283 GGAAGCGGACATGCCCTGCTGAGAGCGGAGAGCCGCTGAACCTCCACGCCACGTCAC 342  
QY 66 Val11eSerLeuProGlnProSerAsp11eProProGly1ThrGlyValPhe11eVal 85  
DB 343 ACGGTACCCCTGCCCTGCTCAGAGACCTTCCCCGGGATGCCGTGCTGGTCACT 402  
QY 86 G1YTYrG1YArGAspAspAsnAap 93  
DB 403 GGCTGGGGCGATGTGACAAATGAT 426  
RESULT 27

AAS20785  
ID AAS20785 standard; DNA; 771 BP.  
XX  
XX AAS20785;  
AC  
XX 09-APR-2002 (first entry)  
XX  
XX DNA encoding human beta-II tryptase active site mutant S194A #5.  
DE  
XX  
XX Human; proteolytic tryptase; protease; recombinant beta-II tryptase;  
KW enzyme; mutant; ds.  
XX  
XX Homo sapiens.  
OS Synthetic.  
XX  
XX WO200198470-A2.  
XX  
XX  
XX 27-DEC-2001.  
XX  
XX 20-JUN-2001; 2001WO-US19681.  
XX  
XX 21-JUN-2000; 2000US-0598982.  
XX  
XX (PROM-) PROMEGA CORP.  
XX  
XX Maffit M, Niles AL, Haak-Frendescho M;  
XX  
XX MPI; 2002-114578/15.  
XX  
XX P-PSDB; AAU12019.  
XX  
XX  
XX DNA construct for producing enzymatically-inactive proteolytic  
PT tryptase, comprises DNA sequence encoding proteolytic tryptase having  
PT an active site mutation -  
XX  
XX  
XX Claim 7; Page 110-111; 126pp; English.  
XX  
XX The present invention relates to recombinant human proteolytic  
CC tryptases, active site mutants of these tryptases and the methods for  
CC producing these. The method involves the production of a DNA expression  
CC construct comprising a promoter operably linked to a secretion signal  
CC sequence which is operably linked to a DNA sequence encoding a  
CC proteolytic tryptase with an active site mutation (the construct drives  
CC expression of a mature proteolytic tryptase that lacks enzymatic activity  
CC due to the active site mutation, in hosts transformed to contain the  
CC construct). The method is useful for producing enzymatically-active  
CC beta-II tryptase. The active site mutants of proteolytic tryptase provide  
CC a tool to investigate the structural and functional properties of the  
CC protease and its enzymatic activity, and for modelling studies. The  
CC enzymatically-active, recombinant proteolytic tryptase produced are  
CC useful as an antigen to generate anti-human tryptase antibodies  
CC and in drug screening for compounds which act as tryptase inhibitors,  
CC antagonists, agonists, etc. AAS20775-AAS20790 encode for recombinant  
CC human beta-II tryptase active site mutants.  
XX  
XX  
SQ Sequence 771 BP; 150 A; 256 C; 240 G; 125 T; 0 other;  
XX  
XX  
XX Alignment Scores:  
Pred. No.: 0.051 Length: 771  
Score: 99.00 Matches: 19  
Percent Similarity: 62.50% Conservative: 11  
Best Local Similarity: 39.58% Mismatches: 18  
Query Match: 17.93% Indels: 0  
DB: 24 Gaps: 0  
US-10-020-441-2 (1-106) x AAS20785 (1-771)  
QY 46 GYPhneAsp11eAla1eValMeLeuAlaG1MeCValAsnLeuGlnSerGly1IeArg 65  
DB 283 GGAAGCGGACATGCCCTGCTGAGAGCGGAGAGCCGCTGAACCTCCACGCCACGTCAC 342  
QY 66 Val11eSerLeuProGlnProSerAsp11eProProGly1ThrGlyValPhe11eVal 85  
DB 343 ACGGTACCCCTGCCCTGCTCAGAGACCTTCCCCGGGATGCCGTGCTGGTCACT 402

OY 86 G1YTYG1YVIGAspAspAspAsp 93  
 |||::|||  
 Db 403 GGCTGGGGCGCATGTGACATGAT 426  
 RESULT 28  
 ID AAS20786  
 XX AAS20786 standard; DNA; 771 BP.  
 AC AAS20786;  
 XX  
 DT 09-APR-2002 (first entry)  
 XX  
 DE DNA encoding human beta-II tryptase active site mutant S194A #6.  
 XX  
 KM Human; proteolytic tryptase; protease; recombinant beta-II tryptase;  
 XX enzyme; mutant; ds.  
 OS Homo sapiens.  
 OS Synthetic.  
 XX  
 PN WO200198470-A2.  
 PD  
 PD 27-DEC-2001.  
 PF 20-JUN-2001; 2001WO-US19681.  
 PR 21-JUN-2000; 2000US-0598982.  
 XX  
 PA (PROM-) PROMEGA CORP.  
 XX  
 PI Maffei M, Niles AL, Haak-Frendscho M;  
 DR WPI: 2002-114578/15.  
 DR P-PSDB; AAU12020.  
 XX  
 PT DNA construct for producing enzymatically-inactive proteolytic  
 PT tryptase, comprises DNA sequence encoding proteolytic tryptase having  
 PT an active site mutation -  
 XX  
 PS Claim 7; Page 113-114; 126pp; English.  
 XX  
 CC The present invention relates to recombinant human proteolytic  
 CC tryptases, active site mutants of these tryptases and the methods for  
 CC producing these. The method involves the production of a DNA expression  
 CC construct comprising a promoter operably linked to a secretion signal  
 CC sequence which is operably linked to a DNA sequence encoding a  
 CC proteolytic tryptase with an active site mutation (the construct drives  
 CC expression of a mature proteolytic tryptase that lacks enzymatic activity  
 CC due to the active site mutation, in hosts transformed to contain the  
 CC construct). The method is useful for producing enzymatically-active  
 CC beta-II tryptase. The active site mutants of proteolytic tryptase provide  
 CC a tool to investigate the structural and functional properties of the  
 CC protease and its enzymatic activity, and for modelling studies. The  
 CC enzymatically-active, recombinant proteolytic tryptase produced are  
 CC useful as an antigen to generate anti-human tryptase antibodies  
 CC and in drug screening for compounds that act as tryptase inhibitors,  
 CC antagonists, agonists, etc. AAS20775-AAS20790 encode for recombinant  
 CC human beta-II tryptase active site mutants.  
 XX  
 SQ Sequence 771 BP; 152 A; 256 C; 240 G; 123 T; 0 other;  
 Alignment Scores:  
 Pred. No.: 0.051 Length: 771  
 Score: 99.00 Matches: 11  
 Percent Similarity: 62.50% Conservative: 19  
 Best Local Similarity: 39.58% Mismatches: 18  
 Query Match: 17.93% Indels: 0  
 DB: 24 Gaps: 0  
 US-10-020-441-2 (1-106) x AAS20786 (1-771)

Db 283 GGAGCGGACATCGCCCTGCTGACTGGAGAGGCCGGTGAACGTCTCCACCAGTCAC 342

Gy 66 ValIleSerLeuProGlnProSerAspIleProRobProGlyThryGlyValPheIleVal 85  
::: :::  
343 ACGGCACCCCTGCCGCCCTGCAGAGAACCCTCCCCGGGGATGCGGTCTGGTGCACT 402  
::: :::

Gy 86 GilyTyrgIyArgAspAapAanAap 93  
::: :::  
Db 403 GGCTGGGGCGATGTGTGACATCATGAT 426  
::: :::

RESULT 29  
AAV44329  
ID AAV44329 standard; DNA, 1137 BP.  
XX AC AAV44329;  
XX DT 24-NOV-1998 (first entry)  
XX DE Human mast cell tryptase I nucleic acid sequence.  
XX KW Mast cell tryptase I; human; MCP-7; mast cell protease 7;  
XX blood clot; anticoagulant; myocardial infarction; reocclusion;  
XX thromboembolism; cerebral embolism; thrombosis; therapy; ss.  
OS Homo sapiens.  
XX Key Location/Qualifiers  
FH CDS 3..824  
FT /\*tag= a  
FN MO9824886-A1.  
PN 11-JUN-1998.  
PD 25-NOV-1997; 97MO-US21620.  
PE 04-DEC-1996; 96US-0032354.  
PR (BGHM ) BRIGHAM & WOMENS HOSPITAL.  
PA Stevens RL;  
PI WPI; 1998-333308/29.  
XX DR P-PSDB; AAW64238.  
XX PT New compositions containing tryptase-7, e.g. mouse mast cell  
PT protease-7 - are used to treat clot formation in e.g. myocardial  
PT infarction, reocclusion following angioplasty or pulmonary  
PT thromboembolism  
XX PS Disclosure; Page 64-65; 92pp; English.

CC This nucleotide sequence includes a coding region for human  
CC mast cell tryptase I (see AAW64238). The invention provides:  
CC compositions comprising an isolated tryptase-7 that may include  
CC chimeric proteins that contain (a) a human tryptase for all but  
CC the active site region and (b) the substrate-binding pocket of  
CC mouse tryptase-7 or its homologues (see AAW64233-39); a method for  
CC treating a blood clot by administering a nucleic acid molecule that  
CC codes for a tryptase-7, or an expression product, to decrease  
CC fibrinogen activity; a nucleic acid encoding a serine protease  
CC (SP), and a method of producing a mature SP by expressing the  
CC inactive zymogen in a host cell, and cleaving the enterokinase  
CC susceptibility domain. The tryptase-7 polypeptides can be used to  
CC treat disorders mediated by undesirable thrombus clot formation  
CC such as myocardial infarction and reocclusion following angioplasty  
CC of blood clots associated with pulmonary thromboembolism, deep vein  
CC thrombosis, cerebral embolism, renal vein and peripheral arterial  
CC thrombosis. They are also useful for all surgical procedures that  
CC require decreased blood clots.

Sequence 1137 BP; 196 A; 410 C; 328 G; 203 T; 0 other;



CC The present invention describes low adenosine (A) content antisense  
CC oligonucleotides and compositions (I) comprising them. In the antisense  
CC oligonucleotides the A is replaced by a 'universal' or alternative base.  
CC (I) can have respiratory, bronchodilator, antiinflammatory, analgesic,  
CC immunosuppressive, antiallergic, hypotensive and cytostatic activities.  
CC The antisense oligonucleotides and (I) can be used to down-regulate the  
CC expression and/or activity of target polypeptides associated with  
CC lung/respiratory disorders and malignancies, such as stimulating and  
CC activating peptide factors and transmitters, transcription factors,  
CC immunoglobulin and antibodies, antibody receptors, cytokines and  
CC chemokines, endogenously produced specific and non-specific enzymes,  
CC binding proteins, adhesion molecules and their receptors, cytokine and  
CC chemokine receptors, adenosine receptors, bradykinin receptors, central  
CC nervous system (CNS) and peripheral nervous and non-nervous system  
CC receptors, CNS and peripheral nervous and non-nervous system peptide  
CC transmitters, defensive, growth factors, vasoactive peptides and  
CC receptors, binding proteins and malignancy associated proteins. The  
CC antisense oligonucleotides may be used in this way to treat disorders  
CC including respiratory obstruction (especially pulmonary obstruction  
CC and/or bronchoconstriction) and/or lung inflammation, allergy(ies)  
CC and/or surfactant hypoproduction which are associated with a disease or  
CC condition selected from pulmonary vasoconstriction, inflammation,  
CC allergies, asthma, impeded respiration, respiratory distress syndrome  
CC (RDS), pain, cystic fibrosis (CF), allergic rhinitis (AR), pulmonary  
CC hypertension, emphysema, chronic obstructive pulmonary disease (COPD),  
CC pulmonary transplantation rejection, pulmonary infections, bronchitis,  
CC and/or cancer. AAF18434 to AAF21543 represent human polynucleotide  
CC fragments and antisense oligonucleotides used in the exemplification of  
CC the present invention.

XX SQ Sequence 1137 BP; 196 A; 410 C; 328 G; 203 T; 0 other;

Alignment Scores:

Pred. No.: 0.0832 Length: 1137  
Score: 99.00 Matches: 19  
Percent Similarity: 62.50% Conservative: 11  
Best Local Similarity: 39.58% Mismatches: 18  
Query Match: 17.93% Indels: 0  
DB: Gaps: 21

US-10-020-441-2 (1-106) x AAF21079 (1-1137)

QY 46 G|YPheasp|leAl|leValMe|leuVala|n|leu|n|se|rg|yl|leArg 65

DB 351 GAGCGGCAATGCGCTGTGAGCGAGGAGCGCGTGAACGTCTCCAGCAGTCCAC 410

QY 66 Val|leSer|leuP|roG|n|ProSe|raSp|leP|roP|roG|l|Y|Th|e|l|Val|Phe|leVal 85

DB 411 ACGGTACCCCTGCCCCCTGCTCAGAGACCTTCCCCCGGAGATGCCGTGGGTACT 470

QY 86 G|YTyrG|lYArG|aSp|aP|aSp|aP|aSp 93

DB 471 GGCTGGGGCGATGTGACAAATGAT 494

RESULT 32 AAA34957

ID AAA34957 standard; DNA; 1137 BP.

AC AAA34957;

XX 28-JUN-2000 (first entry)

XX Human adenosine receptor related polynucleotide SEQ ID NO:2646.

XX Human, adenosine receptor, low adenosine antisense oligonucleotide;  
XX phosphorothiolate, impaired respiration; inflammation; allergy;  
XX allergic disease; bronchoconstriction; inhibitor; antiinflammatory;  
XX antiallergic; antiaesthetic; cytostatic; analgesic; impaired airway;  
XX lung disease; ischaemic condition; pulmonary vasoconstriction; asthma;  
XX respiratory distress syndrome; pain; cystic fibrosis; emphysema;  
XX pulmonary hypertension; chronic obstructive pulmonary disease; COPD;  
XX cancer; leukemia; lymphoma; carcinoma; metastasis; ss.

OS Homo sapiens.

XX PN WO200009525-A2.

XX 24-FEB-2000.

XX 03-AUG-1999; 99WO-US17712.

XX 03-AUG-1998; 98US-0095212.

XX (UYEC-) UNIV EAST CAROLINA.

XX Nyce JW;

XX WPI, 2000-205971/18.

PT New antisense oligonucleotides useful for treating e.g. pulmonary  
PT vasoconstriction, inflammation, allergies, asthma, hypertension,  
PT bronchitis, emphysema, respiratory distress syndrome, ischemia or  
PT cancers -

PS Disclosure; Page 807-808; 1343pp; English.

XX The present invention describes a new composition comprising an  
CC antisense oligonucleotide (ON) with low adenosine (up to 15%), which  
CC targets nucleic acids involved in bronchoconstriction, allergies, and/or  
CC inflammation. The ON can have antiinflammatory, antiallergic,  
CC antiaesthetic, cytostatic and analgesic activities. The compositions are  
CC useful for the treatment of diseases associated with inflammation,  
CC impaired airways, including lung disease and diseases whose secondary  
CC effects afflict the lungs of a subject. They can be used for treating  
CC e.g. ischaemic conditions, pulmonary vasoconstriction, allergies,  
CC asthma, impeded respiration, respiratory distress syndrome, pain, cystic  
CC fibrosis, pulmonary hypertension, emphysema, chronic obstructive  
CC pulmonary disease (COPD), and cancers such as leukemia, lymphomas,  
CC carcinomas, and cancers which may metastasize to the lungs, including  
CC breast and prostate cancer. The reduction of the adenosine content of  
CC the ON reduces side effects. The A-containing ONs break down with the  
CC release of deoxyadenosine which activates adenosine receptors causing  
CC bronchoconstriction and inflammation. AAA3313 to AAA3512 represent the  
CC nucleotide sequences given in the sequence listing from the present  
CC invention, which correspond to SEQ ID NO:1 to 2815, and then the last  
CC 185 sequences are also called SEQ ID NO:1 to 185, but the sequences  
CC differ from the previously named sequences. SEQ ID NO:11 to 1680  
CC (AAA33323 to AAA33992) are specifically claimed ONs from the present  
CC invention. N.B. Sequences given in the disclosure are from the present  
CC invention do not match up with their corresponding SEQ ID NO: sequences  
CC given in the sequence listing.

XX SQ Sequence 1137 BP; 196 A; 410 C; 328 G; 203 T; 0 other;

Alignment Scores:

Pred. No.: 0.0832 Length: 1137  
Score: 99.00 Matches: 19  
Percent Similarity: 62.50% Conservative: 11  
Best Local Similarity: 39.58% Mismatches: 18  
Query Match: 17.93% Indels: 0  
DB: Gaps: 21

US-10-020-441-2 (1-106) x AAA34957 (1-1137)

QY 46 G|YPheasp|leAl|leValMe|leuVala|n|leu|n|se|rg|yl|leArg 65

DB 351 GAGCGGCAATGCGCTGTGAGCGAGGAGCGCGTGAACGTCTCCAGCAGTCCAC 410

QY 66 Val|leSer|leuP|roG|n|ProSe|raSp|leP|roP|roG|l|Y|Th|e|l|Val|Phe|leVal 85

DB 411 ACGGTACCCCTGCCCCCTGCTCAGAGACCTTCCCCCGGAGATGCCGTGGGTACT 470

QY 86 G|YTyrG|lYArG|aSp|aP|aSp|aP|aSp 93

DB 471 GGCTGGGGCGATGTGACAAATGAT 494

RESULT 33  
AAE21076  
ID AAE21076 standard; DNA, 1167 BP.  
XX  
XX AC  
XX AAE21076;  
XX  
XX DT 14-MAR-2001 (first entry)  
XX  
XX DE Human low adenosine antisense oligonucleotide related sequence #2643.  
XX  
XX KW Low adenosine antisense oligonucleotide; phosphorothioate; allergy;  
KW human; airway disorder; bronchoconstriction; lung inflammation;  
KW surfactant depletion; respiratory; bronchodilator; antiinflammatory;  
KW immunosuppressive; antiasthmatic; analgesic; hypotensive; cytosolic;  
KW respiratory obstruction; pulmonary obstruction; impeded respiration;  
KW surfactant hypoproduction; pulmonary vasoconstriction; asthma; RDS;  
KW respiratory distress syndrome; pain; cystic fibrosis; allergic rhinitis;  
KW pulmonary hypertension; emphysema; pulmonary transplantation rejection;  
KW chronic obstructive pulmonary disease; pulmonary infection; bronchitis;  
KW cancer; ss.  
XX  
XX OS Homo sapiens.  
XX  
XX PN WO20062736-A2.  
XX  
XX PD 26-OCT-2000.  
XX  
XX PF 24-MAR-2000; 2000WO-US08020.  
XX  
XX PR 06-APR-1999; 99US-0127958.  
XX  
XX PA (UYEC-) UNIV EAST CAROLINA.  
XX (NYCE/) NYCE J W.  
XX  
XX PI Nyce JW;  
XX  
XX DR WPI; 2000-679539/66.  
XX  
XX PT Low adenosine (A) content antisense oligonucleotides which do not  
XX trigger adenosine receptors during metabolism, useful e.g. for treating  
XX cancers and respiratory obstructions -  
XX  
XX PT Disclosure; Page 878-879; 1592pp; English.  
XX  
XX PS The present invention describes low adenosine (A) content antisense  
XX oligonucleotides and compositions (I) comprising them. In the antisense  
XX oligonucleotides the A is replaced by a 'Universal' or alternative base  
XX (I) can have respiratory, bronchodilator, antiinflammatory, analgesic,  
XX immunosuppressive, antiasthmatic, hypotensive and cytosolic activities  
XX The antisense oligonucleotides and (I) can be used to down-regulate the  
XX expression and or activity of target polypeptides associated with  
XX lung/respiratory disorders and malignancies, such as stimulating and  
XX activating peptide factors and transmitters, transcription factors,  
XX immunoglobulins and antibodies, antibody receptors, cytokines and  
XX chemokines, endogenously produced specific and non-specific enzymes,  
XX binding proteins, adhesion molecules and their receptors, cytokine and  
XX chemokine receptors, adenosine receptors, bradykinin receptors, central  
XX nervous system (CNS) and peripheral nervous and non-nervous system  
XX receptors, CNS and peripheral nervous and non-nervous system peptide  
XX transmitters, defensins, growth factors, vasoactive peptides and  
XX receptors, binding proteins and malignancy associated proteins. The  
XX antisense oligonucleotides may be used in this way to treat disorders  
XX including respiratory obstruction (especially pulmonary obstruction  
XX and/or bronchoconstriction) and/or lung inflammation, allergy(ies)  
XX and/or surfactant hypoproduction which are associated with a disease or  
XX condition selected from pulmonary vasoconstriction, inflammation,  
XX allergies, asthma, impeded respiration, respiratory distress syndrome  
XX (RDS), pain, cystic fibrosis (CF), allergic rhinitis (AR), pulmonary  
XX hypertension, emphysema, chronic obstructive pulmonary disease (COPD),  
XX pulmonary transplantation rejection, pulmonary infections, bronchitis,  
XX and/or cancer. AAE18434 to AAE21543 represent human polynucleotide  
XX fragments and antisense oligonucleotides used in the exemplification of  
XX the present invention.

XX	Sequence 1167 BP; 187 A; 419 C; 342 G; 219 T; 0 other;
SO	
XX	Alignment Scores:
XX	Pred. No.: 0.086 Length: 1167
XX	Score: 99.00 Matches: 20
XX	Percent Similarity: 54.39% Conservative: 11
XX	Best Local Similarity: 35.09% Mismatches: 26
XX	Query Match: 17.93% Indels: 0
XX	DB: 21 Gaps: 0
XX	US-10-020-441-2 (1-106) x AAF21076 (1-1167)
OY	46 GlypheaspilleaIleValMetIleuValIasMetValIasMetGlnSerGlyIleArg 65
DB	7 GAGACGAGCATCGCCCTGCTGAGCTGAGAGAGACCCGCTAAGCTCTCCAGCATCCAC 66
OY	66 ValIleSerIleuPProGlnPProSerAspIleProProGlyIThrGlyValaPheIleVal 85
DB	67 ACGGTACACCTGCCCCCTGCTGCTAGAGACCTTCCCCCGGAGTGGCCGTCGTGGGTCACT 126
OY	86 GlyTYrGlyAArgAspAsnaAspArgAspProSerArgLyshsngLYGly 102
DB	127 GGTCTGGGGCGATGTGACAAATGATGTGGGTCTGGGGAGACGTGAGTGGG 177
XX	RESULT 34
XX	AAA34954
XX	ID AAA34954 standard; DNA; 1167 BP.
XX	AAA34954;
XX	AC
XX	DT 28-JUL-2000 (first entry)
DE	Human adenosine receptor related polynucleotide SEQ ID NO:2643.
XX	Human; adenosine receptor; low adenosine antisense oligonucleotide;
KM	phosphorothioate; impaired respiration; inflammation; allergy;
KM	allergic disease; bronchoconstriction; inhibitor; antiinflammatory;
KM	antiallergic; antilasthmatic; cyostatic; analgesic; impaired airway;
KM	lung disease; ischaemic condition; pulmonary vasoconstriction; asthma;
KM	respiratory distress syndrome; pain; cystic fibrosis; emphysema;
KM	pulmonary hypertension; chronic obstructive pulmonary disease; COPD;
KM	cancer; leukemia; lymphoma; carcinoma; metastasis; ss.
XX	
XX	Homo sapiens.
OS	
XX	WO200009525-A2.
PN	
XX	24-FEB-2000.
PD	
XX	03-AUG-1999; 99WO-US17712.
PF	
XX	03-AUG-1998; 98US-0095212.
PR	
XX	(UYEC-) UNIV EAST CAROLINA.
XX	
PA	
XX	Nyce JW;
PI	
XX	WPI: 2000-205971/18.
DR	
XX	
XX	New antisense oligonucleotides useful for treating e.g. pulmonary
PT	bronchostriction, inflammation, allergies, asthma, hypertension,
PT	asthma, emphysema, respiratory distress syndrome, ischemia or
PT	cancers -
XX	
PS	Disclosure: Page 806-807; 1343bp: English.
XX	
XX	The present invention describes a new composition comprising an
XX	antisense oligonucleotide (ON) with low adenosine (up to 15%), which
CC	targets nucleic acids involved in bronchoconstriction, allergies, and/or
CC	inflammation. The ON can have antiinflammatory, antiallergic,
CC	antiasthmatic, cyostatic and analgesic activities. The compositions are
CC	useful for the treatment of diseases associated with inflammation,

CC impaired airways, including lung disease and diseases whose secondary  
CC effects afflict the lungs of a subject. They can be used for treating  
CC e.g. ischemic conditions, pulmonary vasoconstriction, allergies,  
CC asthma, impeded respiration, respiratory distress syndrome, pain, cystic  
CC fibrosis, pulmonary hypertension, emphysema, chronic obstructive  
CC pulmonary disease (COPD), and cancers such as leukaemia, lymphomas,  
CC carcinomas, and cancers which may metastasise to the lungs, including  
CC breast and prostate cancer. The reduction of the adenine content of  
CC the ONS reduces side effects. The A-containing ONS break down with the  
CC release of deoxyadenosine which activates adenosine receptors causing the  
CC bronchoconstriction and inflammation. AAA3213 to AAA3532 represent the  
CC nucleotide sequences given in the sequence listing from the present  
CC invention, which correspond to SEQ ID NO:1 to 2815, and then the last  
CC 185 sequences are also called SEQ ID NO:1 to 185, but the sequences  
CC differ from the previously named sequences. SEQ ID NO:11 to 1680  
CC (AAA3223 to AAA3992) are specifically claimed ONS from the present  
CC invention. N.B. Sequences given in the disclosure of the present  
CC invention do not match up with their corresponding SEQ ID NO: sequences  
CC given in the sequence listing.

XX  
XX SQ Sequence 1167 BP; 187 A; 419 C; 342 G; 219 T; 0 other;

Alignment Scores:  
Pred. No.: 0.086 Length: 1167  
Score: 99.00 Matches: 20  
Percent Similarity: 54.39% Conservative: 11  
Best Local Similarity: 35.09% Mismatches: 26  
Query Match: 17.93% Indels: 0  
DB: 21 Gaps: 0

US-10-020-441-2 (1-106) x AAA34954 (1-1167)

Qy 46 GLYPheAspIleAlaIleValMetLeuAlaGlnMetValAsnLeuGlnSerGlyIleArg 65  
Db 7 GGAGGGGACATCGCCCTGCTGGAGCGCGGTAGAGCTCCAGCCACATCGCAC 66  
Qy 66 ValIleSerLeuProGlnProSerAspIleProProGlyIleValPheIleVal 85  
Db 67 ACGGTCACCTGCGCCCGCTCAGACAGACCTTCCCGGGGATGCGGCTGGCTACT 126  
Qy 86 GLYTYRGLYAGAspAspAspAspAspProSerAspGlyValAsnGlyGly 102  
Db 127 GGCTGGGGCGATGTGACATGATGTGGCTGGGAGACATGTGAGATGGG 177

RESULT 35  
AAD23854  
ID AAD23854 standard; cDNA; 2662 BP.  
XX  
XX AAD23854;  
XX  
XX 07-MAR-2002 (first entry)  
XX  
XX Human protease PRTS-13 cDNA.  
XX  
XX Human; protease; PRTS-13; tranquilliser; gene therapy; vaccine; allergy;  
XX infection; dermatitis; arteriosclerosis; rheumatoid arthritis; hepatitis;  
XX atherosclerosis; psoriasis; Alzheimer's disease; mental disorder; cancer;  
XX gastrointestinal disorder; Cushing's syndrome; seizure; glaucoma; stroke;  
XX epithelial disorder; urticaria; anorexia; trauma; eczema; nausea;  
XX hypertension; neurological disorder; Parkinson's disease; drug screening;  
XX cardiac; cell proliferative disorder; multiple sclerosis; osteoporosis;  
XX diabetes mellitus; glomerulonephritis; cardiovascular disorder; anaemia;  
XX autoimmune disorder; inflammatory disorder; myocardial infarction; AIDS;  
XX developmental disorder; reproductive disorder; infertility; diarrhoea;  
XX dementia; acidosis; cataract; gynaecomastia; epilepsy; jaundice; ss.  
XX  
XX Homo sapiens.  
XX  
XX  
XX Key Location/Qualifiers  
XX CDS 587..2662  
XX FT /\*tag= a  
XX FT /\*product= "Human protease PRTS-13 protein"  
XX FT sig\_peptide 587..670

FT FT /\*tag= b  
FT mat\_peptide 671..2659  
FT /\*tag= c  
FT /\*product= "Mature human protease PRTS-13 protein"  
XX  
XX W0200183775-A2.  
XX  
XX  
XX 08-NOV-2001.  
XX  
XX 04-MAY-2001; 2001WO-US14651.  
XX  
XX 04-MAY-2000; 2000US-202082P.  
XX PR 11-MAY-2000; 2000US-203566P.  
XX PR 17-MAY-2000; 2000US-205803P.  
XX PR 25-MAY-2000; 2000US-207477P.  
XX PR 01-JUN-2000; 2000US-209402P.  
XX  
XX (INCY-) INCYTE GENOMICS INC.  
XX  
XX Delegeane AM, Lal P, Hafalia A, Patterson C, Walla NK, Kearney L,  
XX Tribouley CM, Khan FA, Yao MG, Baughn MR, Azimzai Y, Elliott VS,  
XX Nguyen DB, Gandhi AR, Yang J, Hernandez R, Policky JL, Lu DM,  
XX Reddy R, Yue H, Tang YT;  
XX  
XX WPI; 2002-034518/04.  
XX P-PSDB; AAE14348.  
XX  
XX Novel human proteases and polynucleotides encoding the proteases,  
XX useful for treating, diagnosing or preventing cell proliferative,  
XX cardiovascular, autoimmune/inflammatory, neurological and developmental  
XX disorders -  
XX  
XX  
XX Claim 5; Page 149-150; 151pp; English.  
XX  
XX  
XX The invention relates to human proteases (PRTS1-14) and its corresponding  
XX cDNA molecules. Human PRTS and its nucleic acid molecule are useful for  
XX the diagnosis, treatment and prevention of disorders associated with  
XX increased or decreased expression of PRTS. Examples of such disorders  
XX include, cell proliferative disorders (arteriosclerosis, atherosclerosis,  
XX hepatitis, psoriasis and cancers); autoimmune/inflammatory disorders  
XX (AIDS, Addison's disease, allergy, anaemia, asthma, atopic dermatitis,  
XX diabetes mellitus, glomerulonephritis, multiple sclerosis, osteoporosis,  
XX trauma, Grave's disease, rheumatoid arthritis, ulcerative colitis, and  
XX viral, bacterial, fungal, parasitic, protozoal and helminthic  
XX infections); cardiovascular disorders (myocardial infarction, ischaemic  
XX heart disease and hypertension); neurological disorders (epilepsy,  
XX Alzheimer's disease, Pick's disease, Huntington's disease, dementia,  
XX Parkinson's disease, stroke, mental disorders including mood, anxiety  
XX and seasonal affective disorder and prion diseases); gastrointestinal  
XX disorders (Crohn's disease, anorexia, nausea, diarrhoea and jaundice);  
XX epithelial disorders (contact dermatitis, eczema, acne vulgaris,  
XX alopecia, scabies, insect bites and urticaria); reproductive disorder  
XX (infertility); disruption of estrous and menstrual cycle and  
XX gynaecomastia); and developmental disorders (renal tubular acidosis,  
XX Cushing's syndrome, seizure disorders, congenital glaucoma and cataract).  
XX PRTS DNA is also in useful in gene therapy. PRTS and its immunogenic  
XX fragments are useful for screening libraries of compounds in several drug  
XX screening assays. The present sequence is human protease PRTS-13 cDNA.  
XX  
XX  
XX SQ Sequence 2662 BP; 403 A; 903 C; 853 G; 503 T; 0 other;

Alignment Scores:  
Pred. No.: 0.244 Length: 2662  
Score: 99.00 Matches: 19  
Percent Similarity: 62.50% Conservative: 11  
Best Local Similarity: 39.58% Mismatches: 18  
Query Match: 17.93% Indels: 0  
DB: 24 Gaps: 0

US-10-020-441-2 (1-106) x AAD23854 (1-2662)

Qy 46 GLYPheAspIleAlaIleValMetLeuAlaGlnMetValAsnLeuGlnSerGlyIleArg 65  
Db 7 GGAGGGGACATCGCCCTGCTGGAGCGCGGTAGAGCTCCAGCCACATCGCAC 66  
Qy 66 ValIleSerLeuProGlnProSerAspIleProProGlyIleValPheIleVal 85  
Db 67 ACGGTCACCTGCGCCCGCTCAGACAGACCTTCCCGGGGATGCGGCTGGCTACT 126  
Qy 86 GLYTYRGLYAGAspAspAspAspAspProSerAspGlyValAsnGlyGly 102  
Db 127 GGCTGGGGCGATGTGACATGATGTGGCTGGGAGACATGTGAGATGGG 177









GenCore version 5.1.4 p5.4578  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: April 9, 2003, 13:07:49 ; Search time 92.0984 Seconds  
(without alignments)  
237.149 Million cell updates/sec

Title: US-10-020-441-2

Perfect score: 106  
Sequence: 1 VSFILRNQDQGIHQPSG.....YGRDNDRDPKRNKGILKK 106

Scoring table: OLIGO  
Gapop 60.0 , Gapext 60.0

Searched: 671580 seqs, 206047115 residues

Word size : 0

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Listing first 1000 summaries

Database : SPTREMBL\_21:\*

- 1: sp\_archaea:\*
- 2: sp\_bacteria:\*
- 3: sp\_fungi:\*
- 4: sp\_human:\*
- 5: sp\_invertebrate:\*
- 6: sp\_mammal:\*
- 7: sp\_mhc:\*
- 8: sp\_organelle:\*
- 9: sp\_phage:\*
- 10: sp\_plant:\*
- 11: sp\_rodent:\*
- 12: sp\_virus:\*
- 13: sp\_vertebrate:\*
- 14: sp\_unclassified:\*
- 15: sp\_virus:\*
- 16: sp\_bacteriap:\*
- 17: sp\_archaeap:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	35	33.0	274	5	Q26552 echistocoma
2	23	21.7	257	5	Q26553 echistocoma
3	8	7.5	1145	5	Q965N2 caenorhabdi
4	7	6.6	149	5	Q95X61 caenorhabdi
5	7	6.6	150	16	Q97P58 streptococ
6	7	6.6	155	6	Q8WN13 bos taurus
7	7	6.6	221	10	Q9LV05 arabidopsis
8	7	6.6	257	8	Q9TEU0 arabidopsis
9	7	6.6	257	8	Q9TER9 arabidopsis
10	7	6.6	257	8	Q9TER8 arabidopsis
11	7	6.6	257	8	Q9T487 arabidopsis
12	7	6.6	257	16	Q92L02 rhizobium m
13	7	6.6	300	16	Q8Y1K0 rhizobium s
14	7	6.6	316	16	Q98H59 rhizobium s
15	7	6.6	336	16	Q8RFR1 fusobacteri
16	7	6.6	342	13	P79783 gallus gall

17	7	6.6	359	17	Q971Z3 sulfolobus
18	7	6.6	382	2	O52988 escherichia
19	7	6.6	382	2	Q91387 escherichia
20	7	6.6	382	2	Q48437 klebsiella
21	7	6.6	382	2	P94856 klebsiella
22	7	6.6	382	2	Q9F4K3 klebsiella
23	7	6.6	406	16	Q98K89 rhizobium 1
24	7	6.6	451	10	Q9FGZ7 arabidopsis
25	7	6.6	478	5	Q9Y1X5 ephedra f
26	7	6.6	623	3	Q9URL7 candida alb
27	7	6.6	657	10	Q9XE83 scorpius bic
28	7	6.6	686	16	Q9A6Y0 arabidopsis
29	7	6.6	794	17	Q8ZVH4 pyrobaculum
30	7	6.6	838	5	Q9W1S4 diatomophila
31	7	6.6	849	4	Q9NT68 homo sapien
32	7	6.6	922	10	Q41058 mus muscu
33	7	6.6	930	11	Q9JLCO mus muscu
34	7	6.6	1173	11	Q63624 ratu
35	7	6.6	1681	5	Q8T1G8 dictyostel
36	7	6.6	1737	4	Q9ULU2 homo sapien
37	7	6.6	2764	11	Q9WTS5 mus muscu
38	7	6.6	2765	11	Q9RIK2 ratu
39	7	6.6	2802	13	Q9DERS gallu
40	7	5.7	22	2	O85558 chlamydia t
41	6	5.7	27	5	Q26523 saccharomy
42	6	5.7	28	3	Q9Y7W7 kluyveriomy
43	6	5.7	43	16	Q8VXF6 mycobacteri
44	6	5.7	72	16	O8X3F5 escherichia
45	6	5.7	76	5	Q9VAP5 diatomophila
46	6	5.7	77	5	Q20690 caenorhabdi
47	6	5.7	79	2	Q9X8Y2 nltrogen-fl
48	6	5.7	83	16	Q9K5Z4 bacillus ha
49	6	5.7	86	2	Q8VR70 escherichia
50	6	5.7	86	5	Q20691 caenorhabdi
51	6	5.7	89	5	Q25033 haemophilus
52	6	5.7	90	2	Q8VU92 lactobacill
53	6	5.7	92	12	Q8UYL2 wheat dwarf
54	6	5.7	92	12	Q89237 wheat dwarf
55	6	5.7	95	16	Q9JZC2 neisseria m
56	6	5.7	95	16	Q9YUG1 neisseria m
57	6	5.7	95	16	Q9YUG1 pseudomonas
58	6	5.7	96	12	Q84425 parametium
59	6	5.7	96	15	Q910B7 human immun
60	6	5.7	99	4	Q8WVH6 homo sapien
61	6	5.7	99	5	Q9W4M2 diatomophila
62	6	5.7	106	5	Q9U374 caenorhabdi
63	6	5.7	106	16	Q8YNN9 anabaena sp
64	6	5.7	122	12	Q8QJ06 hepaticis c
65	6	5.7	125	12	Q8QYQ3 potato leaf
66	6	5.7	131	16	Q92FX7 rickettsia
67	6	5.7	134	3	Q9P694 neurospora
68	6	5.7	134	10	Q40447 nicotiana g
69	6	5.7	135	4	Q9UDC2 homo sapien
70	6	5.7	136	17	Q28093 archaeoglob
71	6	5.7	141	2	Q99QG7 campylobact
72	6	5.7	141	2	Q9AEJ4 campylobact
73	6	5.7	141	2	Q9AEJ3 campylobact
74	6	5.7	141	2	Q9AEJ0 campylobact
75	6	5.7	141	2	Q9AEI9 campylobact
76	6	5.7	142	13	Q919J2 xenopus lae
77	6	5.7	143	13	Q919J3 xenopus lae
78	6	5.7	144	6	Q29186 sus scrofa
79	6	5.7	145	16	Q9A9Y4 caulobacter
80	6	5.7	147	5	Q9SWJ1 anthronomus
81	6	5.7	147	11	Q921U0 mus muscu
82	6	5.7	148	16	Q8YV3 bruceella me
83	6	5.7	149	10	Q9W7S8 arabidopsis
84	6	5.7	149	16	Q8ZC2 yeastinda pe
85	6	5.7	151	12	Q71224 human para
86	6	5.7	151	12	Q71223 human para
87	6	5.7	151	12	Q71225 human para
88	6	5.7	151	12	Q71227 human para
89	6	5.7	151	12	Q71230 human para

90	6	5.7	153	5	Q9B178	Q9b178 caenorhabdi	163	6	5.7	267	16	Q9KG65	Q9kg65 bacillus ha
91	6	5.7	155	5	Q9B177	Q9b177 caenorhabdi	164	6	5.7	270	2	Q54242	Q54242 streptomyc
92	6	5.7	155	16	Q9BN06	Q9bn06 rhizobium 1	165	6	5.7	270	17	Q28620	Q28620 archaeoglob
93	6	5.7	156	2	Q9LAV4	Q9lav4 thermomonas	166	6	5.7	271	16	Q92AS0	Q92as0 listeria in
94	6	5.7	157	16	Q8VOP5	Q8vop5 ralsstonia s	167	6	5.7	272	2	Q8VP91	Q8vp91 burkholderi
95	6	5.7	158	5	Q20692	Q20692 caenorhabdi	168	6	5.7	275	5	Q9K2B0	Q9k2b0 drosophila
96	6	5.7	159	11	Q9JUM8	Q9jum8 mus musculu	169	6	5.7	275	16	Q8KX34	Q8kx34 ralsstonia s
97	6	5.7	160	97	Q8Z6S4	Q8z6s4 salmonella	170	6	5.7	279	16	Q8YLQ0	Q8ylq0 anabaena sp
98	6	5.7	160	17	Q972G4	Q972g4 sulfolobus	171	6	5.7	285	16	Q05403	Q05403 bacillus su
99	6	5.7	166	10	Q39422	Q39422 beta vulgar	172	6	5.7	287	10	Q9C8C0	Q9c8c0 arabisdopsi
100	6	5.7	166	10	Q8S830	Q8s830 oryza sativ	173	6	5.7	287	16	Q9KZ5	Q9kz5 bacillus ha
101	6	5.7	168	10	Q9ZWM3	Q9zwm3 cucumis sat	174	6	5.7	294	16	Q98G8	Q98g8 rhizobium 1
102	6	5.7	170	4	Q14959	Q14959 homo sapien	175	6	5.7	295	16	Q9PNY5	Q9pny5 campylobact
103	6	5.7	171	16	Q93GN6	Q93gn6 salmonella	176	6	5.7	296	2	Q45828	Q45828 clostridium
104	6	5.7	176	16	Q97FJ8	Q97fj8 clostridium	177	6	5.7	296	5	Q99048	Q99048 trypanosoma
105	6	5.7	177	16	Q9HWF0	Q9hwf0 pseudomonas	178	6	5.7	299	16	Q91OR0	Q91or0 pseudomonas
106	6	5.7	178	2	Q9KX17	Q9kx17 bacteroides	179	6	5.7	300	16	Q984K1	Q984k1 rhizobium 1
107	6	5.7	179	16	Q8YXP6	Q8yxp6 anabaena sp	180	6	5.7	302	16	P74286	P74286 synecocyst
108	6	5.7	180	16	Q9CFE9	Q9cfe9 lactococcus	181	6	5.7	303	16	Q8U962	Q8u962 agrobacteri
109	6	5.7	183	16	Q92TT2	Q92tt2 rhizobium m	182	6	5.7	308	11	Q8VG21	Q8vg21 mus musculu
110	6	5.7	183	16	Q8XHL2	Q8xhl2 clostridium	183	6	5.7	310	11	Q8VFP4	Q8vfp4 mus musculu
111	6	5.7	184	10	Q9LL98	Q9ll98 asarum euro	184	6	5.7	311	16	Q97KL2	Q97kl2 clostridium
112	6	5.7	185	10	Q65131	Q65131 liriodendro	185	6	5.7	311	16	Q51622	Q51622 borrelia bu
113	6	5.7	186	10	Q9LI97	Q9li97 calycanthus	186	6	5.7	312	16	Q8RP07	Q8rp07 fusbacteri
114	6	5.7	186	10	Q65130	Q65130 michelia fi	187	6	5.7	314	10	Q04244	Q04244 arabisdopsi
115	6	5.7	187	16	Q9KB72	Q9kb72 bacillus ha	188	6	5.7	314	10	Q8VXQ9	Q8vxq9 clorella f
116	6	5.7	187	16	Q9YER3	Q9yer3 aeropyrum p	189	6	5.7	314	12	Q9YQX5	Q9yxq5 porcine cit
117	6	5.7	188	16	Q50683	Q50683 borrelia bu	190	6	5.7	314	17	Q973B5	Q97b5 sulfolobus
118	6	5.7	189	10	Q9LLA6	Q9lla6 calycanthus	191	6	5.7	315	16	Q96JF7	Q96jf7 rhizobium 1
119	6	5.7	189	16	Q9CDP4	Q9cdp4 mycobacteri	192	6	5.7	316	2	Q9AE42	Q9ae42 rhizobium 1
120	6	5.7	194	10	Q9LLA4	Q9lla4 chloranthus	193	6	5.7	317	10	Q8VXN6	Q8vxn6 coleochaete
121	6	5.7	194	16	Q99213	Q99213 streptococ	194	6	5.7	318	2	Q9KI94	Q9ki94 bacteroides
122	6	5.7	199	2	Q50181	Q50181 mycobacteri	195	6	5.7	318	5	Q9KHQ2	Q9khq2 bacteroides
123	6	5.7	199	16	Q9XL29	Q9xl29 vibrio chol	196	6	5.7	318	5	Q46019	Q46019 caenorhabdi
124	6	5.7	200	5	Q9N805	Q9n805 leishmania	197	6	5.7	318	10	Q8VX58	Q8vx58 spirogyra s
125	6	5.7	200	17	Q8ZYG4	Q8zyg4 pyrobaculum	198	6	5.7	319	16	Q8UD57	Q8ud57 agrobacteri
126	6	5.7	201	4	Q9UMQ5	Q9umq5 homo sapien	199	6	5.7	321	2	Q8RW69	Q8rw69 bacteroides
127	6	5.7	201	16	Q9A8T6	Q9a8t6 caulobacter	200	6	5.7	321	17	Q97K61	Q97k61 sulfolobus
128	6	5.7	202	10	Q9LAA5	Q9lla5 calycanthus	201	6	5.7	322	12	Q98VG9	Q98vg9 feline infe
129	6	5.7	202	17	Q8ZYP2	Q8zyp2 pyrobaculum	202	6	5.7	323	10	Q96518	Q96518 arabisdopsi
130	6	5.7	203	16	Q9PGI6	Q9pgi6 xylella fas	203	6	5.7	323	17	Q97TJ6	Q97tj6 sulfolobus
131	6	5.7	208	11	Q9CWH9	Q9cwh9 mus musculu	204	6	5.7	325	16	Q8XYG2	Q8xyg2 ralsstonia s
132	6	5.7	210	16	Q07229	Q07229 mycobacteri	205	6	5.7	326	5	Q9VZX6	Q9vzx6 drosophila
133	6	5.7	214	5	Q9NF11	Q9nf11 caenorhabdi	206	6	5.7	328	16	Q9CBN8	Q9cbn8 mycobacteri
134	6	5.7	216	10	Q944I4	Q944i4 arabidopsi	207	6	5.7	329	9	Q64303	Q64303 bacterioph
135	6	5.7	216	10	Q944I4	Q944i4 arabidopsi	208	6	5.7	329	9	Q64307	Q64307 enterobacte
136	6	5.7	217	3	Q9USD1	Q9usd1 schizosacch	209	6	5.7	330	10	Q94H71	Q94h71 oryza sativ
137	6	5.7	221	2	Q8RSY6	Q8rsy6 pseudomonas	210	6	5.7	330	16	Q07763	Q07763 mycobacteri
138	6	5.7	221	16	Q25140	Q25140 helicobacte	211	6	5.7	331	16	Q9AD83	Q9ad83 streptomyc
139	6	5.7	223	10	Q943B7	Q943b7 oryza sativ	212	6	5.7	333	16	Q8ZEG7	Q8zeg7 yersinia pe
140	6	5.7	223	16	Q9ZKD5	Q9zkd5 helicobacte	213	6	5.7	334	5	Q8SOX0	Q8sqx0 encephalito
141	6	5.7	223	16	Q9CHA8	Q9cha8 lactococcus	214	6	5.7	335	16	Q8V8I8	Q8v8i8 anabaena sp
142	6	5.7	227	5	Q97266	Q97266 plasmodium	215	6	5.7	339	11	Q99144	Q99144 mus musculu
143	6	5.7	232	5	P81223	P81223 giardia lam	216	6	5.7	341	5	Q25032	Q25032 haemonchus
144	6	5.7	233	5	Q9VFY6	Q9vfy6 drosophila	217	6	5.7	341	11	Q64468	Q64468 mus musculu
145	6	5.7	233	16	Q9X7N2	Q9x7n2 streptomyc	218	6	5.7	342	17	Q9VAX0	Q9vax0 aeropyrum p
146	6	5.7	234	16	Q55877	Q55877 synecocyst	219	6	5.7	343	16	Q981B6	Q981b6 rhizobium 1
147	6	5.7	237	2	Q9APY1	Q9apy1 pseudomonas	220	6	5.7	343	16	Q56026	Q56026 salmonella
148	6	5.7	238	10	Q8S126	Q8s126 oryza sativ	221	6	5.7	346	5	Q9VCS6	Q9vcs6 drosophila
149	6	5.7	239	8	Q9T243	Q9t243 phytophthor	222	6	5.7	348	2	Q56033	Q56033 salmonella
150	6	5.7	240	2	Q93KX5	Q93kx5 streptomyc	223	6	5.7	349	16	Q8RGW9	Q8rgw9 fusbacteri
151	6	5.7	243	17	Q9REPO	Q9rep0 zymomonas m	224	6	5.7	349	17	Q28037	Q28037 archaeoglob
152	6	5.7	243	12	P95893	P95893 sulfolobus	225	6	5.7	350	2	Q917X1	Q917x1 clostridium
153	6	5.7	244	16	Q8U9E7	Q8u9e7 agrobacteri	226	6	5.7	350	16	Q8XJ19	Q8xj19 clostridium
154	6	5.7	248	5	Q9VIT0	Q9vij0 drosophila	227	6	5.7	353	16	Q9KST4	Q9kst4 vibrio chol
155	6	5.7	251	10	Q9CEK7	Q9cek7 arabidopsi	228	6	5.7	355	11	Q64366	Q64366 mus musculu
156	6	5.7	252	10	Q40701	Q40701 oryza sativ	229	6	5.7	356	3	P79037	P79037 emericella
157	6	5.7	252	16	Q8XE24	Q8xe24 escherichia	230	6	5.7	356	16	Q8YCU9	Q8ygu9 bruceella me
158	6	5.7	253	16	Q8ZFP9	Q8zf9 yersinia pe	231	6	5.7	357	5	Q9GRB0	Q9grb0 halocynthia
159	6	5.7	254	10	Q65103	Q65103 oryza sativ	232	6	5.7	359	11	P70554	P70554 ratus norv
160	6	5.7	259	16	Q8U6I6	Q8u6i6 agrobacteri	233	6	5.7	361	10	Q9FMU1	Q9fmu1 arabidopsi
161	6	5.7	263	5	Q20557	Q20557 caenorhabdi	234	6	5.7	363	2	Q9X2Y1	Q9x2y1 bacillus an
162	6	5.7	263	16	Q31859	Q31859 bacillus su	235	6	5.7	363	16	Q98L03	Q98l03 rhizobium 1

236	6	5.7	364	5	Q8S2U0	Q8zuo drosophila	309	6	5.7	443	10	Q94HP5	Q94hp5 oryza sativ
237	6	5.7	365	10	Q9FJ66	Q9fj66 arabidopsis	310	6	5.7	447	5	O17714	O17714 caenorhabd1
238	6	5.7	366	2	Q9KJ91	Q9kj91 vibrio furn	311	6	5.7	447	16	Q8XIF2	Q8xif2 clostridium
239	6	5.7	368	3	Q9C3Z1	Q9c3z1 cochllobol	312	6	5.7	448	5	O21638	O21638 caenorhabd1
240	6	5.7	369	10	Q9MB57	Q9mb57 chlamydom	313	6	5.7	451	5	O19321	O19321 caenorhabd1
241	6	5.7	372	10	Q9SVY9	Q9svy9 arabidopsis	314	6	5.7	452	10	Q9LE51	Q9le51 oryza sativ
242	6	5.7	376	4	O15466	O15466 homo sapien	315	6	5.7	452	13	Q8UWM5	Q8uwm5 oncorhynch
243	6	5.7	378	2	Q45466	Q45466 bacillus sp	316	6	5.7	453	16	Q9RW05	Q9rw05 streptomyce
244	6	5.7	378	2	P95554	P95554 pseudomonas	317	6	5.7	457	5	O18534	O18534 caenorhabd1
245	6	5.7	379	2	O45467	O45467 bacillus sp	318	6	5.7	454	13	Q9ASW7	Q9asw7 oncorhynch
246	6	5.7	381	16	O67659	O67659 aquifex aeo	319	6	5.7	454	13	Q9OZE8	Q9oze8 oncorhynch
247	6	5.7	381	16	Q9RYH0	Q9ryh0 deinococcus	320	6	5.7	454	13	Q8QGE2	Q8qge2 salmo salar
248	6	5.7	383	10	Q8S820	Q8s820 oryza sativ	321	6	5.7	456	10	Q94B55	Q94b55 arabidopsis
249	6	5.7	385	16	Q8UBH6	Q8ubh6 agrobacteri	322	6	5.7	457	5	O18534	Q18534 caenorhabd1
250	6	5.7	386	6	Q9BGU5	Q9bgus bos taurus	323	6	5.7	458	10	P92980	P92980 arabidopsis
251	6	5.7	386	17	Q8TH92	Q8th92 methanocarc	324	6	5.7	458	10	O48887	O48887 arabidopsis
252	6	5.7	387	2	Q9X9A8	Q9x9a8 streptococ	325	6	5.7	459	2	Q8RT11	Q8rt11 magnetite-c
253	6	5.7	387	10	O22843	O22843 arabidopsis	326	6	5.7	460	17	O59186	O59186 pyrococcus
254	6	5.7	387	16	Q9CM28	Q9cm28 pasteurella	327	6	5.7	461	2	O84917	O84917 chloabacillu
255	6	5.7	389	5	Q9VXB9	Q9vxb9 drosophila	328	6	5.7	461	10	Q9ZP70	Q9zpv70 tritlicum ae
256	6	5.7	390	5	Q9VDO8	Q9vdc8 drosophila	329	6	5.7	461	11	O9DCV1	O9dcv1 mus musculu
257	6	5.7	390	16	O83914	O83914 treponema p	330	6	5.7	462	2	Q8RR71	Q8rr71 pseudomonas
258	6	5.7	390	16	O8YFX4	O8yfx4 bruceella me	331	6	5.7	462	11	Q923F4	Q923f4 mus musculu
259	6	5.7	393	16	Q8UBW8	Q8ubw8 agrobacteri	332	6	5.7	462	16	Q9KPR6	Q9kpr6 vibrio chol
260	6	5.7	394	10	Q9SHS1	Q9shs1 arabidopsis	333	6	5.7	462	17	Q96YX7	Q96yx7 sulfolobus
261	6	5.7	394	11	Q9WMD1	Q9wmd1 cricetus	334	6	5.7	464	10	Q9ZP22	Q9zpv22 brassica ju
262	6	5.7	394	11	Q9ROE0	Q9roe0 rattus norv	335	6	5.7	465	16	O86665	O86665 streptomyce
263	6	5.7	394	11	O86693	O86693 mus musculu	336	6	5.7	466	12	O81016	O81016 human papil
264	6	5.7	395	5	Q9U8R6	Q9u8r6 dugesia jap	337	6	5.7	470	10	Q8W1A1	Q8w1a1 glycine max
265	6	5.7	395	10	Q9C693	Q9c693 arabidopsis	338	6	5.7	471	10	O8S289	O8s289 oryza sativ
266	6	5.7	395	11	Q925B4	Q925b4 rattus norv	339	6	5.7	474	5	Q9VQG7	Q9vgq7 drosophila
267	6	5.7	395	11	Q925B4	Q925b4 rattus norv	340	6	5.7	475	10	Q9AMN6	Q9amv6 oryza sativ
268	6	5.7	397	10	Q9SFD1	Q9sfd1 arabidopsis	341	6	5.7	478	2	O92190	O92190 lactobacillu
269	6	5.7	397	16	O983Z0	O983z0 rhizobium 1	342	6	5.7	479	16	Q92U94	Q92u94 rhizobium m
270	6	5.7	398	10	O9ZT13	Q9zt13 zea mays (m	343	6	5.7	481	10	O49935	O49935 sinapis alb
271	6	5.7	400	10	Q9M347	Q9m347 arabidopsis	344	6	5.7	483	16	Q8X170	Q8x170 clostridium
272	6	5.7	400	16	O8ZEL6	O8zel6 yerania pe	345	6	5.7	489	10	O8ZT56	O8zt56 pyrobaculum
273	6	5.7	401	3	O42903	O42903 schizosacch	346	6	5.7	491	10	O9LNY9	O9lvy9 arabidopsis
274	6	5.7	401	6	O08835	O08835 cercopithec	347	6	5.7	491	10	O8W5D1	O8w5d1 oryza sativ
275	6	5.7	401	13	Q9PRG7	Q9prg7 xenopus lae	348	6	5.7	496	16	Q8YQ39	Q8yq39 anabaena sp
276	6	5.7	401	16	Q9PNZ2	Q9pnz2 campylobact	349	6	5.7	499	5	O17555	O17555 caenorhabd1
277	6	5.7	402	5	O8WQ52	O8wq52 giardia lam	350	6	5.7	500	10	O970P0	O970p0 arabidopsis
278	6	5.7	402	10	Q9ARS6	Q9ars6 solanum tub	351	6	5.7	500	10	O24348	O24348 sorghum bic
279	6	5.7	403	5	O07529	O07529 loligo peal	352	6	5.7	501	8	O9GG29	O9gg29 vallisneria
280	6	5.7	404	10	O9LE84	Q9le84 arabidopsis	353	6	5.7	502	10	O22576	O22576 arabidopsis
281	6	5.7	404	16	Q915N8	Q915n8 pseudomonas	354	6	5.7	502	10	O2FPH1	O2fph1 arabidopsis
282	6	5.7	408	16	O8YSM6	O8ysm6 anabaena sp	355	6	5.7	502	10	O24629	O24629 arabidopsis
283	6	5.7	411	16	Q8Y228	Q8y228 ralsctonia s	356	6	5.7	502	16	Q9RJ20	Q9rj20 streptomyce
284	6	5.7	412	11	P70126	P70126 mus musculu	357	6	5.7	506	3	O13604	O13604 schizosacch
285	6	5.7	412	16	O53655	O53655 mycobacteri	358	6	5.7	507	10	Q9FRW7	Q9frw7 nepenthes a
286	6	5.7	413	10	O86282	O86282 oryza sativ	359	6	5.7	508	10	Q9SH91	Q9sh91 arabidopsis
287	6	5.7	415	11	O08605	O08605 mus musculu	360	6	5.7	509	2	Q9RGK3	Q9rgk3 bacterioides
288	6	5.7	417	5	Q9EBU1	Q9ebu1 homo sapien	361	6	5.7	509	2	O31240	O31240 anabaena sp
289	6	5.7	417	5	P92160	P92160 caenorhabd1	362	6	5.7	509	16	P72957	P72957 synchocyst
290	6	5.7	418	4	Q9H5L2	Q9h5l2 homo sapien	363	6	5.7	509	16	O9K3J2	O9kj2 streptomyce
291	6	5.7	418	11	Q9DBU8	Q9dbu8 mus musculu	364	6	5.7	510	5	O62120	O62120 caenorhabd1
292	6	5.7	419	16	O8UDJ8	O8udj8 agrobacteri	365	6	5.7	511	3	Q96WM6	Q96wm6 schizosacch
293	6	5.7	420	16	O8YMH1	O8ymh1 anabaena sp	366	6	5.7	511	4	Q96D08	Q96d08 homo sapien
294	6	5.7	421	16	Q9ABJ7	Q9abj7 caulobacter	367	6	5.7	511	5	Q9GYG5	Q9gyg5 caenorhabd1
295	6	5.7	422	10	Q38948	Q38948 arabidopsis	368	6	5.7	515	5	O8TIN3	O8tin3 dictyostell
296	6	5.7	422	11	Q8ROE1	O8roe1 mus musculu	369	6	5.7	515	16	O9CJ06	O9cjc6 pasteurella
297	6	5.7	424	4	O9NS35	O9ns35 homo sapien	370	6	5.7	519	2	Q9JPF0	Q9jpf0 neisseria m
298	6	5.7	426	10	Q9ZV30	Q9zv30 arabidopsis	371	6	5.7	519	5	Q9U251	Q9u251 caenorhabd1
299	6	5.7	428	2	Q93AT6	Q93at6 erwinia chr	372	6	5.7	519	10	O46555	O46555 oryza sativ
300	6	5.7	429	5	Q9BMF3	Q9bmf3 manduca sex	373	6	5.7	519	16	Q9JTD2	Q9jtd2 neisseria m
301	6	5.7	430	10	Q8SIW3	Q8siw3 oryza sativ	374	6	5.7	525	11	O70145	O70145 mus musculu
302	6	5.7	431	10	Q9C8D1	Q9c8d1 arabidopsis	375	6	5.7	526	17	Q973V2	Q973v2 sulfolobus
303	6	5.7	433	16	Q8ZRP5	Q8zrp5 salmonella	376	6	5.7	529	16	Q98QX5	Q98qx5 mycoplasma
304	6	5.7	435	10	Q9M6C6	Q9m6c6 arabidopsis	377	6	5.7	533	10	Q8RY80	Q8ry80 arabidopsis
305	6	5.7	438	5	O8T8P6	O8t8p6 locusta mig	378	6	5.7	534	16	Q97PA2	Q97pa2 streptococ
306	6	5.7	439	16	Q91IK7	Q91ik7 pseudomonas	379	6	5.7	537	17	Q97B53	Q97b53 thermoplasma
307	6	5.7	439	16	Q986Z0	Q986z0 rhizobium 1	380	6	5.7	541	16	O8YL89	O8yl89 anabaena sp
308	6	5.7	443	5	Q95QH7	Q95qh7 caenorhabd1	381	6	5.7	544	2	Q9XJR2	Q9xjr2 pseudomonas

382	6	5.7	545	17	026908	026908 methanobact	455	6	5.7	817	10	081512	081512 arabidopsis
383	6	5.7	550	16	08UFT7	08uft7 agrobacteri	456	6	5.7	828	10	064477	064477 arabidopsis
384	6	5.7	555	5	08SYW8	08syw8 dtrosophila	457	6	5.7	833	16	08Ym29	08Ym29 anabaena sp
385	6	5.7	559	2	09FS01	09fs01 anabaena va	458	6	5.7	840	16	09RSJ4	09RSJ4 deinococcus
386	6	5.7	560	3	09P6r8	09p6r8 neurospora	459	6	5.7	841	10	064542	064542 arabidopsis
387	6	5.7	563	10	024596	024596 zea mays (m	460	6	5.7	846	5	0962j7	0962j7 plasmodium
388	6	5.7	563	16	09CL06	09cl06 pasteurella	461	6	5.7	856	16	0927f3	0927f3 listeria in
389	6	5.7	565	12	09JUB49	09je49 guinea pig	462	6	5.7	856	16	08Y320	08Y320 listeria mo
390	6	5.7	567	3	P78871	P78871 eschizosacch	463	6	5.7	865	5	09N998	09N998 leishmania
391	6	5.7	568	12	080T31	080t31 human parai	464	6	5.7	869	2	09F609	09F609 escherichia
392	6	5.7	571	12	09E7N1	09e7n1 human parai	465	6	5.7	869	2	09EYm6	09EYm6 escherichia
393	6	5.7	574	5	017902	017902 caenorhabdi	466	6	5.7	874	10	09FLB4	09FLB4 arabidopsis
394	6	5.7	577	16	097GH4	097gh4 clostridium	467	6	5.7	879	16	08XCP4	08XCP4 escherichia
395	6	5.7	580	12	09E7M9	09e7m9 chimeric hu	468	6	5.7	883	16	08XAX3	08XAX3 escherichia
396	6	5.7	582	16	08RB11	08rb11 thermoaer	469	6	5.7	889	10	09M1M3	09M1M3 arabidopsis
397	6	5.7	587	12	09E7M7	09e7m7 chimeric hu	470	6	5.7	895	16	08EYm3	08EYm3 bruceella me
398	6	5.7	589	5	08T0B2	08t0b2 dtrosophila	471	6	5.7	896	16	098EA3	098ea3 rhizobium 1
399	6	5.7	590	4	09H2P2	09h2p2 homo sapien	472	6	5.7	898	5	09UJ13	09UJ13 caenorhabdi
400	6	5.7	593	10	09SHM0	09shmo arabidopsis	473	6	5.7	899	5	09NFX2	09NFX2 dtrosophila
401	6	5.7	597	16	08YP33	08yp33 anabaena sp	474	6	5.7	899	5	076935	076935 dtrosophila
402	6	5.7	599	10	09LDC1	09ldc1 beta vulgar	475	6	5.7	899	5	09VG23	09VG23 dtrosophila
403	6	5.7	600	2	P72289	P72289 rhizobium 1	476	6	5.7	902	5	09NFX3	09NFX3 dtrosophila
404	6	5.7	608	10	093ZU8	093zu8 arabidopsis	477	6	5.7	902	5	076934	076934 dtrosophila
405	6	5.7	609	13	091776	091776 xenopus lae	478	6	5.7	902	5	09VCV4	09vcv4 dtrosophila
406	6	5.7	611	2	047847	047847 pantoea egg	479	6	5.7	906	16	09RTN7	09rtn7 deinococcus
407	6	5.7	621	5	08S531	08s531 encephalito	480	6	5.7	909	5	09Y007	09Y007 plasmodium
408	6	5.7	624	16	08XW86	08xw86 clostridium	481	6	5.7	910	10	093Y57	093Y57 arabidopsis
409	6	5.7	625	2	030704	030704 mycoplasma	482	6	5.7	914	11	091W15	091w15 mus musculu
410	6	5.7	625	10	09LOV9	09lov9 arabidopsis	483	6	5.7	924	4	015327	015327 homo sapien
411	6	5.7	627	2	030643	030643 mycoplasma	484	6	5.7	931	10	09ZEP3	09ZEP3 arabidopsis
412	6	5.7	631	10	091LT74	091lt74 arabidopsis	485	6	5.7	934	2	09RBB5	09RBB5 escherichia
413	6	5.7	634	5	044903	044903 caenorhabdi	486	6	5.7	934	2	09RBB4	09RBB4 escherichia
414	6	5.7	634	16	069728	069728 mycobacteri	487	6	5.7	934	16	09K452	09K452 streptomyce
415	6	5.7	637	4	09NY28	09ny28 homo sapien	488	6	5.7	935	2	093UJ3	093UJ3 escherichia
416	6	5.7	651	2	09RAI8	09rai8 myxobacter	489	6	5.7	935	2	08VL95	08VL95 escherichia
417	6	5.7	652	15	083364	083364 moloney mur	490	6	5.7	935	2	08VL00	08VL00 escherichia
418	6	5.7	654	15	089683	089683 murine leuk	491	6	5.7	936	2	09LB66	09LB66 escherichia
419	6	5.7	655	4	09NRE0	09nre0 homo sapien	492	6	5.7	936	2	093FK0	093FK0 citrobacter
420	6	5.7	657	16	09PMD2	09pmd2 campylobact	493	6	5.7	937	2	093K94	093K94 escherichia
421	6	5.7	661	4	09H696	09h696 homo sapien	494	6	5.7	938	2	09EXN4	09EXN4 escherichia
422	6	5.7	661	5	09NKE9	09nke9 dtrosophila	495	6	5.7	938	2	052620	052620 escherichia
423	6	5.7	664	5	09VUW2	09vw2 dtrosophila	496	6	5.7	938	2	08VL93	08VL93 escherichia
424	6	5.7	666	16	09X838	09x838 streptomyce	497	6	5.7	939	2	047018	047018 escherichia
425	6	5.7	670	4	09PIS9	09pis9 homo sapien	498	6	5.7	939	2	09RBR5	09RBR5 escherichia
426	6	5.7	670	4	09H347	09h347 homo sapien	499	6	5.7	939	2	09RCH1	09RCH1 escherichia
427	6	5.7	680	4	096S23	096s23 homo sapien	500	6	5.7	939	2	093UJ4	093UJ4 escherichia
428	6	5.7	681	5	095BM4	095bm4 dtrosophila	501	6	5.7	939	2	093K93	093K93 escherichia
429	6	5.7	697	4	094968	094968 homo sapien	502	6	5.7	939	2	09SBR2	09SBR2 escherichia
430	6	5.7	699	13	073735	073735 xenopus lae	503	6	5.7	939	2	P77067	P77067 escherichia
431	6	5.7	706	5	023600	023600 caenorhabdi	504	6	5.7	939	16	08YKT5	08YKT5 anabaena sp
432	6	5.7	708	2	09F5G1	09f5g1 agrobacteri	505	6	5.7	940	2	006896	006896 escherichia
433	6	5.7	710	3	014407	014407 neurospora	506	6	5.7	942	10	09FMJ7	09FMJ7 arabidopsis
434	6	5.7	713	16	08XYT0	08xyt0 ralestonia g	507	6	5.7	946	10	09CSJ3	09CSJ3 arabidopsis
435	6	5.7	714	13	090XD9	090xd9 gallus gall	508	6	5.7	948	2	09RCR3	09RCR3 escherichia
436	6	5.7	717	12	08QYQ6	08qyq6 potato leaf	509	6	5.7	948	2	08RSS9	08RSS9 escherichia
437	6	5.7	732	10	09FHL3	09fhl3 arabidopsis	510	6	5.7	948	2	08RNT8	08RNT8 escherichia
438	6	5.7	743	5	017305	017305 caenorhabdi	511	6	5.7	949	16	09KU47	09KU47 vibrio chol
439	6	5.7	750	2	P70921	P70921 bradyrhizob	512	6	5.7	950	5	020733	020733 caenorhabdi
440	6	5.7	757	2	00S992	00s992 alcaligenes	513	6	5.7	959	4	09UPZ7	09upz7 homo sapien
441	6	5.7	762	2	09E293	09e293 zymomonas m	514	6	5.7	981	12	091LJ7	091lj7 white spot
442	6	5.7	764	4	08WUG7	08wug7 homo sapien	515	6	5.7	985	5	09Y142	09Y142 dtrosophila
443	6	5.7	772	16	08UWX3	08uwx3 agrobacteri	516	6	5.7	985	5	09V4X8	09V4X8 dtrosophila
444	6	5.7	774	3	09P444	09p444 emericella	517	6	5.7	1002	10	09SMW5	09smw5 oryza sativ
445	6	5.7	774	5	08T4G4	08t4g4 dtrosophila	518	6	5.7	1011	5	096S27	096s27 schistosoma
446	6	5.7	776	5	09S7M4	09s7m4 dtrosophila	519	6	5.7	1017	16	006586	006586 mycobacteri
447	6	5.7	785	10	09LVN2	09lvn2 arabidopsis	520	6	5.7	1018	5	09VGP4	09VGP4 dtrosophila
448	6	5.7	790	17	097UN0	097un0 sulfolobus	521	6	5.7	1031	10	094LQ7	094LQ7 oryza sativ
449	6	5.7	791	4	08WXB1	08wxb1 homo sapien	522	6	5.7	1043	17	097AX7	097ax7 thermoplasm
450	6	5.7	793	5	09Y108	09y108 dtrosophila	523	6	5.7	1044	2	00S923	00S923 listeria g
451	6	5.7	800	17	08ZXD1	08zxd1 pyrobaculum	524	6	5.7	1059	10	09FKM9	09fkm9 arabidopsis
452	6	5.7	812	2	09RBA5	09rba5 atrobacte	525	6	5.7	1075	10	094V19	094v19 oryza sativ
453	6	5.7	812	5	021623	021623 caenorhabdi	526	6	5.7	1112	5	095ZL4	095zl4 caenorhabdi
454	6	5.7	815	17	0976T6	0976t6 sulfolobus	527	6	5.7	1121	10	08W302	08w302 oryza sativ

528	6	5.7	1122	13	P70047	P70047 xenopus lae	601	5	4.7	33	15	Q74746	Q74746 human immun
529	6	5.7	1161	10	Q9FP84	Q9IFB4 arabidopsis	602	5	4.7	34	16	Q9T2H1	Q9T2H1 cyanidium c
530	6	5.7	1193	10	Q93W04	Q93W04 oryza sativ	603	5	4.7	34	16	Q8USP3	Q8USP3 agrobacteri
531	6	5.7	1219	10	Q9AY59	Q9AY59 oryza sativ	604	5	4.7	35	16	Q8Z189	Q8Z189 yerania pe
532	6	5.7	1228	5	Q61198	Q61198 caenorhabdi	605	5	4.7	36	8	Q9GFA9	Q9GFA9 acorus cala
533	6	5.7	1229	5	Q9VF61	Q9VF61 drosophila	606	5	4.7	36	8	Q9GFA3	Q9GFA3 cabomba car
534	6	5.7	1245	2	Q9L470	Q9L470 staphylococ	607	5	4.7	36	8	Q9GFP7	Q9GFP7 ceratophyll
535	6	5.7	1249	12	Q6S152	Q6S152 african swi	608	5	4.7	36	8	Q9GFB9	Q9GFB9 dirimys wint
536	6	5.7	1250	16	Q9PPL7	Q9PPL7 campylobact	609	5	4.7	36	8	Q9GFB1	Q9GFB1 gnetum gnet
537	6	5.7	1254	6	Q29122	Q29122 sus scrofa	610	5	4.7	36	8	Q9GF76	Q9GF76 lactoris fe
538	6	5.7	1270	13	Q9PWF5	Q9PWF5 morone saxa	611	5	4.7	36	8	Q9GFT4	Q9GFT4 litidendro
539	6	5.7	1276	13	Q91BD1	Q91BD1 gallus gall	612	5	4.7	36	8	Q9MSR0	Q9MSR0 zania furfu
540	6	5.7	1277	11	Q35821	Q35821 rattus norv	613	5	4.7	36	8	Q9MSP9	Q9MSP9 nymphea od
541	6	5.7	1284	4	Q9HCE6	Q9HCE6 homo sapien	614	5	4.7	37	4	Q96G90	Q96G90 homo sapien
542	6	5.7	1285	4	Q9UEG2	Q9UEG2 homo sapien	615	5	4.7	40	12	Q91WZ3	Q91WZ3 human echov
543	6	5.7	1285	4	Q9BZ27	Q9BZ27 homo sapien	616	5	4.7	40	12	Q91WX8	Q91WX8 human coxa
544	6	5.7	1290	2	Q9RFK9	Q9RFK9 stigmatalia	617	5	4.7	40	12	Q91MX5	Q91MX5 human enter
545	6	5.7	1325	4	Q9CDD1	Q9CDD1 homo sapien	618	5	4.7	41	16	Q9X065	Q9X065 thermotoga
546	6	5.7	1358	13	Q73798	Q73798 xenopus lae	619	5	4.7	42	6	Q8MNX1	Q8MNX1 ovis aries
547	6	5.7	1406	5	Q9V4D4	Q9V4D4 drosophila	620	5	4.7	43	4	Q9UC08	Q9UC08 homo sapien
548	6	5.7	1406	5	Q9XIU0	Q9XIU0 drosophila	621	5	4.7	43	6	Q46662	Q46662 macropus ro
549	6	5.7	1408	5	Q9XZU7	Q9XZU7 drosophila	622	5	4.7	43	6	Q9URP7	Q9URP7 coprinus ci
550	6	5.7	1460	11	Q8R4S1	Q8R4S1 mus musculu	623	5	4.7	44	2	Q8RQK8	Q8RQK8 bacillus ci
551	6	5.7	1501	11	Q8R4P9	Q8R4P9 mus musculu	624	5	4.7	46	2	Q8RQK8	Q8RQK8 bacillus ci
552	6	5.7	1513	4	Q9NXY3	Q9NXY3 homo sapien	625	5	4.7	48	16	Q8UBD4	Q8UBD4 agrobacteri
553	6	5.7	1620	4	Q9ULTR	Q9ULTR homo sapien	626	5	4.7	49	2	Q82954	Q82954 ralteonla s
554	6	5.7	1634	3	Q96TY0	Q96TY0 neurospora	627	5	4.7	49	2	Q50263	Q50263 phytoplasma
555	6	5.7	1688	5	Q8SXB0	Q8SXB0 drosophila	628	5	4.7	49	12	Q9NAN6	Q9NAN6 hepaticis c
556	6	5.7	1789	5	Q8T2E3	Q8T2E3 dictyosteli	629	5	4.7	50	12	Q9YTV4	Q9YTV4 human echov
557	6	5.7	1955	10	Q8W078	Q8W078 oryza sativ	630	5	4.7	50	12	Q91FE2	Q91FE2 chilo iride
558	6	5.7	2061	4	Q9BYW2	Q9BYW2 homo sapien	631	5	4.7	51	16	Q9Z858	Q9Z858 chlamydia p
559	6	5.7	2111	3	Q9C102	Q9C102 schizosacch	632	5	4.7	51	12	Q66826	Q66826 human echov
560	6	5.7	2161	12	Q91724	Q91724 grapevine r	633	5	4.7	53	2	Q8ROK2	Q8ROK2 bacillus ce
561	6	5.7	2161	12	Q91901	Q91901 rupestris s	634	5	4.7	53	16	Q92ZK1	Q92ZK1 rhizobium m
562	6	5.7	2165	12	Q91RW8	Q91RW8 human respi	635	5	4.7	54	2	Q9RG92	Q9RG92 borrelia he
563	6	5.7	2165	12	Q41356	Q41356 human respi	636	5	4.7	55	16	Q8RAZ9	Q8RAZ9 thermoaer
564	6	5.7	2165	12	P90197	P90197 human respi	637	5	4.7	55	17	Q971A1	Q971A1 sulfolobus
565	6	5.7	2165	12	Q82021	Q82021 human respi	638	5	4.7	56	2	Q93KE7	Q93KE7 chlamydia p
566	6	5.7	2165	12	Q41355	Q41355 human respi	639	5	4.7	56	10	Q9XPT4	Q9XPT4 chlamydomon
567	6	5.7	2166	12	Q36C35	Q36C35 human respi	640	5	4.7	57	2	P71460	P71460 lactobacilli
568	6	5.7	2166	12	Q36C37	Q36C37 human respi	641	5	4.7	58	8	Q94VR1	Q94VR1 fletcherod
569	6	5.7	2245	5	Q94378	Q94378 caenorhabdi	642	5	4.7	58	15	Q87599	Q87599 chimpanzee
570	6	5.7	2301	5	Q95ZD0	Q95ZD0 leishmania	643	5	4.7	58	16	Q9WYH6	Q9WYH6 thermotoga
571	6	5.7	2340	16	Q9ZDP1	Q9ZDP1 rickettsia	644	5	4.7	60	17	Q9HRS3	Q9HRS3 halobacteri
572	6	5.7	2382	4	Q9H4A3	Q9H4A3 homo sapien	645	5	4.7	61	3	Q9USR9	Q9USR9 schizosacch
573	6	5.7	2391	10	Q9XE40	Q9XE40 oryza sativ	646	5	4.7	61	5	Q95ZM3	Q95ZM3 caenorhabdi
574	6	5.7	2646	12	Q9WPZ7	Q9WPZ7 gill-associ	647	5	4.7	61	17	Q980Q6	Q980Q6 sulfolobus
575	6	5.7	2652	2	Q9RA21	Q9RA21 vibrio mari	648	5	4.7	62	2	Q48650	Q48650 lactococcus
576	6	5.7	2703	5	Q9VEG7	Q9VEG7 drosophila	649	5	4.7	62	12	Q8VAT3	Q8VAT3 white spot
577	6	5.7	2715	5	Q61603	Q61603 cercophila	650	5	4.7	62	16	Q8YXK2	Q8YXK2 aibaena sp
578	6	5.7	3105	12	Q8UZX9	Q8UZX9 cerophthic	651	5	4.7	62	17	Q8ZTE4	Q8ZTE4 pyrobaculum
579	6	5.7	3389	4	Q96QU9	Q96QU9 homo sapien	652	5	4.7	63	16	Q8XZJ9	Q8XZJ9 escherichia
580	6	5.7	3391	12	Q9WDAS	Q9WDAS dengue vtru	653	5	4.7	64	12	Q999V5	Q999V5 human coxa
581	6	5.7	3391	12	Q9WDAS	Q9WDAS dengue vtru	654	5	4.7	64	16	Q9A136	Q9A136 streptococ
582	6	5.7	3508	4	Q96RM4	Q96RM4 homo sapien	655	5	4.7	65	10	Q9LE28	Q9LE28 vicia faba
583	6	5.7	3868	5	Q9W090	Q9W090 drosophila	656	5	4.7	65	10	Q944E8	Q944E8 oryza sativ
584	6	5.7	4085	12	Q9DLN0	Q9DLN0 human coron	657	5	4.7	65	16	Q8YPA8	Q8YPA8 anaena sp
585	6	5.7	4203	5	Q96SG2	Q96SG2 caenorhabdi	658	5	4.7	66	2	Q8RLS0	Q8RLS0 providencia
586	6	5.7	4215	5	Q9W332	Q9W332 drosophila	659	5	4.7	66	5	Q94780	Q94780 trypanosoma
587	6	5.7	4219	5	Q9NLE7	Q9NLE7 caenorhabdi	660	5	4.7	67	15	Q87608	Q87608 chimpanzee
588	6	5.7	6758	12	Q9DLN1	Q9DLN1 human coron	661	5	4.7	67	16	Q31586	Q31586 bacillus ru
589	5	4.7	10	10	Q08938	Q08938 nicotiana t	662	5	4.7	68	6	Q9XBZ2	Q9XBZ2 bacillus ce
590	5	4.7	21	2	Q93S20	Q93S20 staphylococ	663	5	4.7	68	6	P79312	P79312 sus scrofa
591	5	4.7	21	13	Q9PRQ1	Q9PRQ1 oncorhynch	664	5	4.7	68	16	Q9CGB6	Q9CGB6 lactococcus
592	5	4.7	24	11	Q88570	Q88570 mus musculu	665	5	4.7	68	16	Q8ZKX3	Q8ZKX3 salmonella
593	5	4.7	24	11	Q88571	Q88571 mus spretus	666	5	4.7	68	16	Q9XKZ1	Q9XKZ1 streptomyce
594	5	4.7	24	11	Q9CW21	Q9CW21 rattus sp.	667	5	4.7	69	7	Q31360	Q31360 brachydanio
595	5	4.7	27	5	Q963K6	Q963K6 nephasoma m	668	5	4.7	69	9	Q9ZXU9	Q9ZXU9 bacteriopho
596	5	4.7	27	12	P89660	P89660 hepatitis c	669	5	4.7	69	16	Q8ZYE6	Q8ZYE6 rhizobium m
597	5	4.7	27	12	Q91L56	Q91L56 hepatitis c	670	5	4.7	69	16	Q8ZXR0	Q8ZXR0 yersinia pe
598	5	4.7	31	4	Q9UD38	Q9UD38 homo sapien	671	5	4.7	70	4	Q9UKN6	Q9UKN6 homo sapien
599	5	4.7	31	6	Q9N1P8	Q9N1P8 macropus eu	672	5	4.7	70	12	Q66477	Q66477 human enter
600	5	4.7	32	8	Q8SL89	Q8SL89 euglena ste	673	5	4.7	70	15	Q87601	Q87601 chimpanzee

674	5	4.7	70	15	087613	087613 chimpanzee	747	5	4.7	87	2	09R605	09R605 escherichia
675	5	4.7	70	15	087620	087620 chimpanzee	748	5	4.7	87	2	09R604	09R604 yersinia en
676	5	4.7	71	2	093KE8	093KE8 chlamydia p	749	5	4.7	87	13	09PVR8	09PVR8 oryzae lat
677	5	4.7	71	7	095588	095588 brachydanio	750	5	4.7	87	16	09PHK1	09PHK1 xyella fas
678	5	4.7	71	10	0857G9	0857G9 oryza sativ	751	5	4.7	87	16	09CLV3	09CLV3 pasteurilla
679	5	4.7	71	16	082BX6	082BX6 yersinia pe	752	5	4.7	87	16	P71603	P71603 mycobacteri
680	5	4.7	72	2	09RGA4	09RGA4 borrelia he	753	5	4.7	88	2	09R690	09R690 borrelia he
681	5	4.7	72	15	087610	087610 chimpanzee	754	5	4.7	88	2	054516	054516 streptococ
682	5	4.7	72	16	092FU9	092FU9 listeria in	755	5	4.7	88	10	093YK9	093YK9 braasica na
683	5	4.7	72	16	091093	091093 streptomyce	756	5	4.7	88	11	08VCU9	08VCU9 mus musculu
684	5	4.7	73	2	0915R7	0915R7 salmonella	757	5	4.7	88	16	09PD18	09PD18 xyella fas
685	5	4.7	73	8	09XQR3	09XQR3 piscum sativ	758	5	4.7	88	16	09ZP04	09ZP04 streptococ
686	5	4.7	73	8	09BBQ7	09BBQ7 lotus japon	759	5	4.7	88	16	08TXA1	08TXA1 anabaena sp
687	5	4.7	73	11	070134	070134 mus musculu	760	5	4.7	88	17	09VOR0	09VOR0 pyrococcus
688	5	4.7	73	12	09Q2Q9	09Q2Q9 human papil	761	5	4.7	89	2	P70925	P70925 bacillus 11
689	5	4.7	73	12	09DMW4	09DMW4 human papil	762	5	4.7	89	5	P91660	P91660 drosophila
690	5	4.7	73	15	087595	087595 chimpanzee	763	5	4.7	89	5	09VWH5	09VWH5 drosophila
691	5	4.7	73	16	09P9X6	09P9X6 xyella fas	764	5	4.7	89	7	031355	031355 brachydanio
692	5	4.7	73	16	08YAV9	08YAV9 listeria mo	765	5	4.7	89	15	088103	088103 chimpanzee
693	5	4.7	74	16	09RUB2	09RUB2 deinococcus	766	5	4.7	89	17	08U1Q3	08U1Q3 pyrococcus
694	5	4.7	74	17	09HQU2	09HQU2 halobacteri	767	5	4.7	89	17	08TUQ8	08TUQ8 methanosarc
695	5	4.7	75	8	09MTN4	09MTN4 oenothera h	768	5	4.7	90	2	09RGA6	09RGA6 borrelia he
696	5	4.7	75	9	021995	021995 streptococ	769	5	4.7	90	2	09RGA2	09RGA2 borrelia he
697	5	4.7	75	15	087606	087606 chimpanzee	770	5	4.7	90	5	09BMO8	09BMO8 bombyx mori
698	5	4.7	75	16	09KIL9	09KIL9 neisseria m	771	5	4.7	90	8	08WGB7	08WGB7 procambarus
699	5	4.7	76	16	09WY01	09WY01 thermotoga	772	5	4.7	90	9	08SCH9	08SCH9 propionibac
700	5	4.7	77	5	08T6F0	08T6F0 drosophila	773	5	4.7	90	10	09SNG0	09SNG0 oryza sativ
701	5	4.7	77	5	08STA2	08STA2 drosophila	774	5	4.7	90	16	092VT2	092VT2 rhizobium m
702	5	4.7	77	10	094DX6	094DX6 oryza sativ	775	5	4.7	90	16	08RHM1	08RHM1 fusobacteri
703	5	4.7	77	12	08V2L8	08V2L8 camelipox vi	776	5	4.7	91	5	09XYH9	09XYH9 leishmania
704	5	4.7	77	16	09KMI0	09KMI0 vibrio chol	777	5	4.7	91	10	09LWE6	09LWE6 oryza sativ
705	5	4.7	77	16	09I3T0	09I3T0 pseudomonas	778	5	4.7	91	11	09JLJ1	09JLJ1 mus musculu
706	5	4.7	77	16	0982I9	0982I9 rhizobium 1	779	5	4.7	91	16	08U550	08U550 agrobacteri
707	5	4.7	78	2	09WTH1	09WTH1 escherichia	780	5	4.7	92	2	09S1H1	09S1H1 staphylococ
708	5	4.7	78	16	097RD4	097RD4 streptococ	781	5	4.7	92	2	0933I3	0933I3 chlamydia p
709	5	4.7	78	16	08YR29	08YR29 anabaena sp	782	5	4.7	92	10	09L1S4	09L1S4 oryza sativ
710	5	4.7	79	4	09UDF0	09UDF0 homo sapien	783	5	4.7	92	15	087484	087484 chimpanzee
711	5	4.7	79	5	095SD5	095SD5 drosophila	784	5	4.7	92	16	09JRF3	09JRF3 neisseria m
712	5	4.7	79	16	08XTX3	08XTX3 ralstonia s	785	5	4.7	93	2	093EF8	093EF8 helicobacte
713	5	4.7	80	2	085366	085366 enterococcu	786	5	4.7	93	7	031260	031260 rattus norv
714	5	4.7	80	2	094076	094076 mycoplasma	787	5	4.7	93	16	09KOT5	09KOT5 vibrio chol
715	5	4.7	80	10	09FWI7	09FWI7 oryza sativ	788	5	4.7	94	3	09UT11	09UT11 schizosach
716	5	4.7	80	11	09CV90	09CV90 mus musculu	789	5	4.7	94	12	091LD5	091LD5 white spot
717	5	4.7	80	16	025814	025814 helicobacte	790	5	4.7	94	16	0971Y6	0971Y6 clostridium
718	5	4.7	80	16	08XBR5	08XBR5 escherichia	791	5	4.7	95	2	0936W3	0936W3 pseudomonas
719	5	4.7	80	16	08R768	08R768 thermoaer	792	5	4.7	95	2	093EE2	093EE2 helicobacte
720	5	4.7	81	4	09H7M3	09H7M3 homo sapien	793	5	4.7	95	5	044033	044033 plasmodium
721	5	4.7	81	4	096906	096906 homo sapien	794	5	4.7	95	5	044038	044038 plasmodium
722	5	4.7	81	5	095SF3	095SF3 drosophila	795	5	4.7	95	8	09TEVO	09TEVO opsanus tau
723	5	4.7	81	8	09TKW7	09TKW7 nephrosemi	796	5	4.7	95	11	091YAO	091YAO mus musculu
724	5	4.7	82	2	08RJ30	08RJ30 chlamydia p	797	5	4.7	95	16	09K1F7	09K1F7 neisseria m
725	5	4.7	82	4	09H396	09H396 homo sapien	798	5	4.7	95	16	08YPG7	08YPG7 anabaena sp
726	5	4.7	82	4	096BPI	096BPI homo sapien	799	5	4.7	95	17	08TYR2	08TYR2 methanosarc
727	5	4.7	82	12	09Q3Q6	09Q3Q6 hepatitis c	800	5	4.7	96	2	007160	007160 mycobacteri
728	5	4.7	82	16	08YH97	08YH97 brucella me	801	5	4.7	96	12	09QCVR	09QCVR tomato leaf
729	5	4.7	83	2	087578	087578 rhizobium 1	802	5	4.7	96	15	09ICE9	09ICE9 human immun
730	5	4.7	83	5	017929	017929 caenorhabdi	803	5	4.7	96	15	09IOB6	09IOB6 human immun
731	5	4.7	83	16	08ZEA7	08ZEA7 yersinia pe	804	5	4.7	96	15	09IOB9	09IOB9 human immun
732	5	4.7	83	16	08XP54	08XP54 clostridium	805	5	4.7	96	16	08XN05	08XN05 clostridium
733	5	4.7	84	9	09AZI9	09AZI9 bacterioph	806	5	4.7	97	5	095UI2	095UI2 drosophila
734	5	4.7	84	16	09JVG8	09JVG8 neisseria m	807	5	4.7	97	6	09TSB6	09TSB6 trichosurus
735	5	4.7	84	16	09CUG3	09CUG3 lactococcus	808	5	4.7	97	10	086EB4	086EB4 oryza sativ
736	5	4.7	84	17	08U106	08U106 pyrococcus	809	5	4.7	97	16	09CPM4	09CPM4 lactococcus
737	5	4.7	85	2	0933S2	0933S2 chlamydia p	810	5	4.7	98	2	09RGR2	09RGR2 borrelia he
738	5	4.7	85	10	08W3B8	08W3B8 oryza sativ	811	5	4.7	98	9	08SC59	08SC59 brx2 conyer
739	5	4.7	85	11	063058	063058 rattus norv	812	5	4.7	98	10	P93137	P93137 flagellaria
740	5	4.7	85	12	09YML5	09YML5 lymantria d	813	5	4.7	98	10	08RVV5	08RVV5 hellanhus
741	5	4.7	85	16	P72576	P72576 synecocyst	814	5	4.7	98	16	09AIX2	09AIX2 streptococ
742	5	4.7	85	16	007207	007207 mycobacteri	815	5	4.7	98	16	097SV3	097SV3 streptococ
743	5	4.7	86	4	014603	014603 homo sapien	816	5	4.7	98	16	097FD1	097FD1 clostridium
744	5	4.7	86	4	09UDU0	09UDU0 homo sapien	817	5	4.7	98	16	08X4B3	08X4B3 escherichia
745	5	4.7	86	10	041737	041737 zea mays (m	818	5	4.7	99	8	P92508	P92508 ampullaria
746	5	4.7	86	16	0910E9	0910E9 streptomyce	819	5	4.7	99	15	09WQR8	09WQR8 human immun





```

966 5 4.7 117 10 Q42090 Q42090 arabidopsis
967 5 4.7 117 10 Q9LM29 Q9LM29 arabidopsis
968 5 4.7 117 10 Q9LGA2 Q9LGA2 oryza sativ
969 5 4.7 117 10 Q94287 Q94287 oryza sativ
970 5 4.7 117 10 Q9SJ28 Q9SJ28 arabidopsis
971 5 4.7 117 12 Q65584 Q65584 bovine herp
972 5 4.7 117 16 Q9K039 Q9K039 neisseria m
973 5 4.7 117 16 Q9JVK4 Q9JVK4 neisseria m
974 5 4.7 118 2 P72075 P72075 neisseria g
975 5 4.7 118 2 Q93Q18 Q93Q18 aeromonas s
976 5 4.7 118 4 Q43722 Q43722 homo sapien
977 5 4.7 118 8 Q94V68 Q94V68 gila cypha
978 5 4.7 118 10 Q42184 Q42184 arabidopsis
979 5 4.7 118 10 Q9F1T2 Q9F1T2 arabidopsis
980 5 4.7 118 10 Q93XE4 Q93XE4 oryza sativ
981 5 4.7 118 11 Q921M1 Q921M1 mus musculu
982 5 4.7 118 11 Q9D1H8 Q9D1H8 mus musculu
983 5 4.7 118 16 Q8VK13 Q8VK13 mycobacteri
984 5 4.7 119 3 Q9C1N0 Q9C1N0 penicillium
985 5 4.7 119 4 Q9UBJ9 Q9UBJ9 homo sapien
986 5 4.7 119 9 Q8SDV7 Q8SDV7 bacterioph
987 5 4.7 119 11 Q9WV58 Q9WV58 mus musculu
988 5 4.7 119 16 Q8UEU0 Q8UEU0 agrobacteri
989 5 4.7 119 16 Q9RTT1 Q9RTT1 deinococcus
990 5 4.7 119 17 Q57831 Q57831 pyrococcus
991 5 4.7 120 12 Q68991 Q68991 turkey aden
992 5 4.7 120 16 Q98D60 Q98D60 rhizobium 1
993 5 4.7 120 16 Q930Y6 Q930Y6 rhizobium m
994 5 4.7 120 17 Q8TUG8 Q8TUG8 bacillus su
995 5 4.7 120 17 Q8TUG8 Q8TUG8 methanosarc
996 5 4.7 121 2 Q8RTU2 Q8RTU2 xanthomonas
997 5 4.7 121 4 Q96LM4 Q96LM4 homo sapien
998 5 4.7 121 5 Q9GSB6 Q9GSB6 bombyx mori
999 5 4.7 121 6 Q9N0A2 Q9N0A2 macaca fasc
1000 5 4.7 121 8 Q48208 Q48208 drosophila

```

## ALIGNMENTS

## RESULT 1

```

Q26552 PRELIMINARY; PRT; 274 AA.
ID Q26552
AC Q26552;
DT 01-NOV-1996 (TREMBlrel. 01, Created)
DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)
DT 01-MAR-2002 (TREMBlrel. 20, Last annotation update)
DE Elastase.
OS Schistosoma mansoni (Blood fluke).
OC Eukaryota; Metazoa; Platyhelminthes; Trematoda; Digenea; Strigeidida;
OC Schistosomatidae; Schistosomatidae; Schistosoma.
OX NCBI_Taxid=6183;
[1]
RN
RP SEQUENCE FROM N.A.
RC STRAIN=PUERTO RICO;
RX MEDLINE=96362066; PubMed=8720180;
RA Pierrot C., Capron A., Khalife J.;
RT "Cloning and characterization of two genes encoding Schistosoma
mansoni elastase."
RL Mol. Biochem. Parasitol. 75:113-117(1995).
CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1, ALSO KNOWN AS THE
CC TRYPSIN FAMILY.
DR EMBL: U31768; AAC46967.1; -.
DR HSSP: P20231; IAAO.
DR MEROPS: S01.144; -.
DR InterPro: IPR001314; Chymotrypsin.
DR InterPro: IPR001254; Ser_protease_Try.
DR Pfam: PF00089; trypsin; 1.
DR PRINTS: PR00722; CHYMOTRYPSIN.
DR SMART: SM00020; TRYPSIN_DOM; 1.
DR PROSITE: PS50240; TRYPSIN_DOM; 1.
DR PROSITE: PS00134; TRYPSIN_HIS; UNKNOWN_1.
KW Hydrolyase; Serine protease.

```

SO SEQUENCE 274 AA; 29586 MW; 53F21914FDDDA450 CRC64;

Query Match 33.0%; Score 35; DB 5; Length 274;  
 Best Local Similarity 100.0%; Pred. No. 7.4e-29;  
 Matches 35; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 VSFLTRNGDQGGIHQPSGVKVAPEGYPSCMSAR 35  
 |||||  
 Db 84 VSFLTRNGDQGGIHQPSGVKVAPEGYPSCMSAR 118

## RESULT 2

```

Q26553 PRELIMINARY; PRT; 257 AA.
ID Q26553
AC Q26553;
DT 01-NOV-1996 (TREMBlrel. 01, Created)
DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)
DT 01-MAR-2002 (TREMBlrel. 20, Last annotation update)
DE Elastase.
OS Schistosoma mansoni (Blood fluke).
OC Eukaryota; Metazoa; Platyhelminthes; Trematoda; Digenea; Strigeidida;
OC Schistosomatidae; Schistosomatidae; Schistosoma.
OX NCBI_Taxid=6183;
[1]
RN
RP SEQUENCE FROM N.A.
RC STRAIN=PUERTO RICO;
RX MEDLINE=96362066; PubMed=8720180;
RA Pierrot C., Capron A., Khalife J.;
RT "Cloning and characterization of two genes encoding Schistosoma
mansoni elastase."
RL Mol. Biochem. Parasitol. 75:113-117(1995).
CC EMBL: U31769; AAC46968.1; -.
DR HSSP: P00766; IGCD.
DR InterPro: IPR001254; Ser_protease_Try.
DR Pfam: PF00089; trypsin; 1.
DR SMART: SM00020; TRYPSIN_DOM; 1.
DR PROSITE: PS50240; TRYPSIN_DOM; 1.
DR PROSITE: PS00134; TRYPSIN_HIS; UNKNOWN_1.
KW Hydrolyase; Serine protease.
SO SEQUENCE 257 AA; 27971 MW; 00F887279F9D64 CRC64;

```

Query Match 21.7%; Score 23; DB 5; Length 257;  
 Best Local Similarity 100.0%; Pred. No. 3.9e-16;  
 Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 3 FLTLRNGDQGGIHQPSGVKVAPE 25  
 |||||  
 Db 82 FLTLRNGDQGGIHQPSGVKVAPE 104

## RESULT 3

```

Q965N2 PRELIMINARY; PRT; 1145 AA.
ID Q965N2
AC Q965N2;
DT 01-DEC-2001 (TREMBlrel. 19, Created)
DT 01-DEC-2001 (TREMBlrel. 19, Last sequence update)
DT 01-MAR-2002 (TREMBlrel. 20, Last annotation update)
DE Hypothetical protein BE003N10.3.
GN BE003N10.3.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditidae;
OC Rhabditidae; Pelodicerinae; Caenorhabditis.
OX NCBI_Taxid=6239;
[1]
RN
RP SEQUENCE FROM N.A.
RC STRAIN=BRISTOL N2;
RX MEDLINE=99069613; PubMed=9851916;
RA None;
RT "Genome sequence of the nematode C. elegans: a platform for
investigating biology. The C. elegans Sequencing Consortium."
RL Science 282:2012-2018(1998).
RN [2]
RP SEQUENCE FROM N.A.

```

RC STRAIN=BRISTOL N2;  
 RA Waterston R.;  
 RT "Direct Submission."  
 RL Submitted (JUN-2001) to the EMBL/GenBank/DBJ databases.  
 DR EMBL, ACC92660; AAK73857.1;  
 DR InterPro: IPR002485; DUF13.  
 DR InterPro: IPR003006; IG\_MHC.  
 DR Pfam: PF01482; DUF13; 1.  
 DR Prodom: PD002726; DUF13; 1.  
 DR PROSITE: PS00290; IG\_MHC; UNKNOWN\_1.  
 SQ SEQUENCE 1145 AA; 128815 MW; 67EC2437F4FA377 CRC64;

Query Match  
 Best Local Similarity 100.0%; Pred. No. 11;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 59 NLOSGIRV 66  
 |||||  
 Db 708 NLOSGIRV 715

## RESULT 4

Q95X61 PRELIMINARY; PRT; 149 AA.  
 ID Q95X61  
 AC Q95X61;  
 DT 01-DEC-2001 (TREMBlrel. 19, Created)  
 DT 01-DEC-2001 (TREMBlrel. 19, Last sequence update)  
 DT 01-MAR-2002 (TREMBlrel. 20, Last annotation update)  
 DE Hypothetical 16.1 KDa protein.  
 GN K11D12.11.  
 OS Caenorhabditis elegans.  
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditidae;  
 OC Rhabditidae; Peloderinae; Caenorhabditis.  
 NCBI\_TaxID=6239;

RA [1]  
 RC SEQUENCE FROM N.A.  
 RP STRAIN=BRISTOL N2;  
 RX MEDLINE=99069613; PubMed=9851916;  
 RA None;

RT "Genome sequence of the nematode C. elegans: a platform for  
 RT investigating biology. The C. elegans Sequencing Consortium.";  
 RL Science 282:2012-2018 (1998).  
 RN [2]

RP SEQUENCE FROM N.A.  
 RC STRAIN=BRISTOL N2;  
 RA Henkhaus J., Wohlmann P., Giliam B.;  
 RT "The sequence of C. elegans cosmid K11D12."  
 RL Submitted (JAN-1998) to the EMBL/GenBank/DBJ databases.  
 RN [3]

RP SEQUENCE FROM N.A.  
 RC STRAIN=BRISTOL N2;  
 RA Waterston R.;  
 RT "Direct Submission."  
 RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.  
 DR EMBL, AF039047; AAK95882.1;  
 DR InterPro: IPR002223; Kunitz\_BPTI.  
 DR Pfam: PF00014; Kunitz\_BPTI; 1.  
 DR Prodom: PD000222; Kunitz\_BPTI; 1.  
 DR PROSITE: PS00280; BPTI\_KUNITZ\_1; UNKNOWN\_1.  
 DR PROSITE: PS50279; BPTI\_KUNITZ\_2; 1.  
 KM Hypothetical protein; Serine protease inhibitor.  
 SQ SEQUENCE 149 AA; 16138 MW; 32C6D4CDD5FE7F20 CRC64;

Query Match  
 Best Local Similarity 100.0%; Pred. No. 24;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 43 TLGFDI 49  
 |||||  
 Db 111 TLGFDI 117

## RESULT 5

Q97P58 PRELIMINARY; PRT; 150 AA.  
 ID Q97P58  
 AC Q97P58;  
 DT 01-OCT-2001 (TREMBlrel. 18, Created)  
 DT 01-OCT-2001 (TREMBlrel. 18, Last sequence update)  
 DT 01-MAR-2002 (TREMBlrel. 20, Last annotation update)  
 DE Hypothetical protein SP1786.  
 GN SP1786.  
 OS Streptococcus pneumoniae.  
 OC Bacteria; Firmicutes; Bacillus/Clostridium group; Lactobacillales;  
 OC Streptococcaceae; Streptococcus.  
 NCBI\_TaxID=1313;  
 RN [1]

RP SEQUENCE FROM N.A.  
 RC STRAIN=TIGR4;  
 RX MEDLINE=21357209; PubMed=11463916;  
 RA Tectelin H., Nelson K.E., Paulsen I.T., Eisen J.A., Read T.D.,  
 RA Peterson S., Heidelberg J., DeBoy R.T., Haft D.H., Dodson R.J.,  
 RA Durkin A.S., Gwinn M., Kolonay J.F., Nelson W.C., Peterson J.D.,  
 RA Mayhew L.A., White O., Salzberg S.L., Lewis M.R., Radune D.,  
 RA Holtzapple E., Khouri H., Wolf A.M., Uffnerback T.R., Hansen C.L.,  
 RA McDonald L.A., Feldblyum T.V., Angiuoli S., Dickinson T., Hickey E.K.,  
 RA Holt I.E., Loftus B.J., Yang F., Smith H.O., Venter J.C.,  
 RA Dougherty B.A., Morrison D.A., Hollingshead S.K., Fraser C.M.;  
 RT "Complete genome sequence of a virulent isolate of Streptococcus  
 RT pneumoniae."  
 RL Science 293:498-506 (2001).  
 DR EMBL, AE007472; AAK75859.1;  
 DR TIGR, SP1786;  
 DR InterPro: IPR02114; HPr\_Serp\_site.  
 DR PROSITE: PS00589; PTS\_HPR\_SER; UNKNOWN\_1.  
 KM Hypothetical protein; Complete proteome.  
 SQ SEQUENCE 150 AA; 16659 MW; D6FADF8A5BA442C CRC64;

Query Match  
 Best Local Similarity 100.0%; Pred. No. 24;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 69 LPOPSDI 75  
 |||||  
 Db 63 LPOPSDI 69

## RESULT 6

Q8WN13 PRELIMINARY; PRT; 165 AA.  
 ID Q8WN13  
 AC Q8WN13;  
 DT 01-MAR-2002 (TREMBlrel. 20, Created)  
 DT 01-MAR-2002 (TREMBlrel. 20, Last sequence update)  
 DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)  
 DE Sialyltransferase 8 (Fragment).  
 OS Bos taurus (Bovine).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;  
 OC Bovidae; Bovinae; Bos.  
 NCBI\_TaxID=9913;  
 RN [1]

RP SEQUENCE FROM N.A.  
 RA Natalizio A., Ruggiero D., Wiernsperger N., Lagarde M.;  
 RT "Implication of biosynthetic enzymes in modifications of  
 RT glycosphingolipids induced by AGEs in retinal pericytes."  
 RL Submitted (NOV-2001) to the EMBL/GenBank/DBJ databases.  
 DR EMBL, AF450088; AAL47018.1;  
 DR InterPro: IPR001675; GT\_29.  
 DR Pfam: PF00777; Glyco\_transf\_29; 1.  
 KM Transferrase.

FT NON\_TER 1  
 FT NON\_TER 1  
 SQ SEQUENCE 165 AA; 18513 MW; 5ABC3370A11F1B95 CRC64;

Query Match  
 Best Local Similarity 100.0%; Pred. No. 26;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 100 NGGILKK 106  
| | | | |  
Db 12 NGGILKK 18

## RESULT 7

Q9LVU5 PRELIMINARY; PRT; 221 AA.  
ID Q9LVU5;  
AC Q9LVU5;  
DT 01-OCT-2000 (TREMBlrel. 15, Created)  
DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)  
DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)  
DE Similarity to ribosomal protein L9.  
OS Arabidopsis thaliana (mouse-ear cress).  
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
OC Spermatophyta; Magnoliopsida; eudicotyledons; core eudicots; Rosidae;  
OC eustosids 11; Brassicales; Brassicaceae; Arabidopsis.  
OX NCBI\_TaxID=3702;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=COLUMBIA;  
RX MEDLINE=20181125; PubMed=10718197;  
RA Sato S., Nakamura Y., Kaneo T., Katoh T., Asamizu E., Kotani H.,  
RA Tabata S.;  
RT "Structural analysis of Arabidopsis thaliana chromosome 5. X. Sequence  
RT features of the regions of 3,076,755 bp covered by sixty P1 and TAC  
RT clones.";  
RL DNA Res. 7:31-63(2000).  
DR EMBL; AB018116; BAA87147.1; -;  
DR InterPro; IPR000244; Ribosomal\_L9.  
DR Pfam; PF01281; Ribosomal\_L9; 1.  
KW Ribosomal protein.  
SQ SEQUENCE 221 AA; 25351 MW; EBB37A89C7BD6524 CRC64;

Query Match 6.6%; Score 7; DB 10; Length 221;  
Best Local Similarity 100.0%; Pred. No. 34;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 21 VKVARGY 27  
| | | | |  
Db 62 VKVARGY 68

## RESULT 8

Q9TEUO PRELIMINARY; PRT; 257 AA.  
ID Q9TEUO;  
AC Q9TEUO;  
DT 01-MAY-2000 (TREMBlrel. 13, Created)  
DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)  
DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)  
DE Cytochrome b (Fragment).  
GN CYTB.  
OS Apalone mutica.  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Testudines; Cryptodira; Trionychidae; Trionychidae; Apalone.  
OX NCBI\_TaxID=101490;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=FIER-M;  
RX MEDLINE=20098405; PubMed=10631049;  
RA Weisrock D.W., Janzen F.J.;  
RT "Comparative molecular phylogeography of North American softshell  
RT turtles (Apalone): implications for regional and wide-scale historical  
RT evolutionary forces.";  
RL Mol. Phylogenet. Evol. 14:152-164(2000).  
CC -1- FUNCTION: COMPONENT OF THE UBIQUINOL-CYTOCHROME C REDUCTASE  
CC COMPLEX (COMPLEX III OR CYTOCHROME B-C1 COMPLEX), WHICH IS A  
CC RESPIRATORY CHAIN THAT GENERATES AN ELECTROCHEMICAL POTENTIAL  
CC COUPLED TO ATP SYNTHESIS (BY SIMILARITY).  
CC -1- COFACTOR: TWO HEME GROUPS (B562 AND B566) WHICH ARE NOT COVALENTLY  
CC BOUND TO THE PROTEIN (BY SIMILARITY).

CC -1- SUBUNIT: THE MAIN SUBUNITS OF COMPLEX B-C1 ARE: CYTOCHROME B,  
CC CYTOCHROME C1 AND THE RISKE PROTEIN (BY SIMILARITY).  
CC -1- SIMILARITY: BELONGS TO THE CYTOCHROME B FAMILY.

DR EMBL; AF168764; AAD51310.1; -;  
DR InterPro; IPR000179; Cyt b b6.  
DR Pfam; PF00032; cytochrome\_b\_c1.  
DR Pfam; PF00033; cytochrome\_b\_n; 1.  
DR PROSITE; PS00193; CYTOCHROME\_B\_OO; UNKNOWN 1.  
KW Electron transport; Heme; Mitochondrion; Respiratory chain;  
KW Transmembrane.  
FT NON TER  
SQ SEQUENCE 257 AA; 28904 MW; 23A5A95F734B84D CRC64;

Query Match 6.6%; Score 7; DB 8; Length 257;  
Best Local Similarity 100.0%; Pred. No. 38;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 38 RPIAQT 44  
| | | | |  
Db 197 RPIAQT 203

## RESULT 9

Q9TE9 PRELIMINARY; PRT; 257 AA.  
ID Q9TE9;  
AC Q9TE9;  
DT 01-MAY-2000 (TREMBlrel. 13, Created)  
DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)  
DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)  
DE Cytochrome b (Fragment).  
GN CYTB.  
OS Apalone mutica.  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Testudines; Cryptodira; Trionychidae; Trionychidae; Apalone.  
OX NCBI\_TaxID=101490;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=LACR1-M;  
RX MEDLINE=20098405; PubMed=10631049;  
RA Weisrock D.W., Janzen F.J.;  
RT "Comparative molecular phylogeography of North American softshell  
RT turtles (Apalone): implications for regional and wide-scale historical  
RT evolutionary forces.";  
RL Mol. Phylogenet. Evol. 14:152-164(2000).  
CC -1- FUNCTION: COMPONENT OF THE UBIQUINOL-CYTOCHROME C REDUCTASE  
CC COMPLEX (COMPLEX III OR CYTOCHROME B-C1 COMPLEX), WHICH IS A  
CC RESPIRATORY CHAIN THAT GENERATES AN ELECTROCHEMICAL POTENTIAL  
CC COUPLED TO ATP SYNTHESIS (BY SIMILARITY).  
CC -1- COFACTOR: TWO HEME GROUPS (B562 AND B566) WHICH ARE NOT COVALENTLY  
CC BOUND TO THE PROTEIN (BY SIMILARITY).  
CC -1- SUBUNIT: THE MAIN SUBUNITS OF COMPLEX B-C1 ARE: CYTOCHROME B,  
CC CYTOCHROME C1 AND THE RISKE PROTEIN (BY SIMILARITY).  
CC -1- SIMILARITY: BELONGS TO THE CYTOCHROME B FAMILY.  
DR EMBL; AF168766; AAD51312.1; -;  
DR InterPro; IPR000179; Cyt b b6.  
DR Pfam; PF00032; cytochrome\_b\_c1.  
DR Pfam; PF00033; cytochrome\_b\_n; 1.  
DR PROSITE; PS00193; CYTOCHROME\_B\_OO; UNKNOWN 1.  
KW Electron transport; Heme; Mitochondrion; Respiratory chain;  
KW Transmembrane.  
FT NON TER  
SQ SEQUENCE 257 AA; 28889 MW; 324073A5F734B84D CRC64;

Query Match 6.6%; Score 7; DB 8; Length 257;  
Best Local Similarity 100.0%; Pred. No. 38;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 38 RPIAQT 44  
| | | | |  
Db 197 RPIAQT 203

```
RESULT 10
Q9TE8      PRELIMINARY;      PRT;      257 AA.
ID Q9TE8
AC Q9TE8;
DT 01-MAY-2000 (TREMBlrel. 13, Created)
DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)
DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
DE Cytochrome b (Fragment).
GN CYTB.
OS Apalone mutica.
OC Mitochondrion.
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Testudines; Cryptodira; Trionychidae; Trionychidae; Apalone.
OX NCBI_TaxID=101490;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=TXR;
RX MEDLINE=20098405; PubMed=10631049;
RA Weisrock D.W., Janzen F.J.;
RT "Comparative molecular phylogeography of North American softshell
turtles (Apalone): implications for regional and wide-scale historical
evolutionary forces."
RL Mol. Phylogenet. Evol. 14:152-164(2000).
CC -1- FUNCTION: COMPONENT OF THE UBIQUINOL-CYTOCHROME C REDUCTASE
CC COMPLEX (COMPLEX III OR CYTOCHROME B-C1 COMPLEX), WHICH IS A
CC RESPIRATORY CHAIN THAT GENERATES AN ELECTROCHEMICAL POTENTIAL
CC COUPLED TO ATP SYNTHESIS (BY SIMILARITY).
CC -1- COFACTOR: TWO HEME GROUPS (B562 AND B566) WHICH ARE NOT COVALENTLY
CC BOUND TO THE PROTEIN (BY SIMILARITY).
CC -1- SUBUNIT: THE MAIN SUBUNITS OF COMPLEX B-C1 ARE: CYTOCHROME B,
CC CYTOCHROME C1 AND THE RISKE PROTEIN (BY SIMILARITY).
DR EMBL; AF168767; AAD51313.1; -.
DR InterPro: IPR0010179; Cyt_b_b6.
DR Pfam; PF00032; cytochrome_b_c1.
DR Pfam; PF00033; cytochrome_b_n; 1.
DR PROSITE; PS00193; CYTOCHROME_B_QO; UNKNOWN_1.
KM Electron transport; Heme; Mitochondrion; Respiratory chain;
KW Transmembrane.
FT NON_TER
SQ SEQUENCE 257 AA; 28931 MW; 0EA34C4FF5132AA8 CRC64;

Query Match      6.6%; Score 7; DB 8; Length 257;
Best Local Similarity 100.0%; Pred. No. 38;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 38 RPIAQT 44
Db 197 RPIAQT 203

RESULT 11
Q9T4S7      PRELIMINARY;      PRT;      257 AA.
ID Q9T4S7
AC Q9T4S7;
DT 01-MAY-2000 (TREMBlrel. 13, Created)
DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)
DT 01-DEC-2001 (TREMBlrel. 18, Last annotation update)
DE Cytochrome b (Fragment).
GN CYTB.
OS Apalone mutica.
OC Mitochondrion.
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Testudines; Cryptodira; Trionychidae; Trionychidae; Apalone.
OX NCBI_TaxID=101490;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=IACR1-M, AND ARWR1;
RA Weisrock D.W., Janzen F.J.;
RT "Comparative molecular phylogeography of North American softshell
turtles (Apalone): implications for regional and wide-scale historical
evolutionary forces."
RL Mol. Phylogenet. Evol. 0:0-0(1999).
```

```
CC -1- FUNCTION: COMPONENT OF THE UBIQUINOL-CYTOCHROME C REDUCTASE
CC COMPLEX (COMPLEX III OR CYTOCHROME B-C1 COMPLEX), WHICH IS A
CC RESPIRATORY CHAIN THAT GENERATES AN ELECTROCHEMICAL POTENTIAL
CC COUPLED TO ATP SYNTHESIS (BY SIMILARITY).
CC -1- COFACTOR: TWO HEME GROUPS (B562 AND B566) WHICH ARE NOT COVALENTLY
CC BOUND TO THE PROTEIN (BY SIMILARITY).
CC -1- SUBUNIT: THE MAIN SUBUNITS OF COMPLEX B-C1 ARE: CYTOCHROME B,
CC CYTOCHROME C1 AND THE RISKE PROTEIN (BY SIMILARITY).
DR EMBL; AF168765; AAD51311.1; -.
DR EMBL; AF168763; AAD51309.1; -.
DR InterPro: IPR0010179; Cyt_b_b6.
DR Pfam; PF00032; cytochrome_b_c1.
DR Pfam; PF00033; cytochrome_b_n; 1.
DR PROSITE; PS00193; CYTOCHROME_B_QO; UNKNOWN_1.
KM Electron transport; Heme; Mitochondrion; Respiratory chain;
KW Transmembrane.
FT NON_TER
SQ SEQUENCE 257 AA; 28931 MW; 0EB1C4D5F7287C5A CRC64;

Query Match      6.6%; Score 7; DB 8; Length 257;
Best Local Similarity 100.0%; Pred. No. 38;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 38 RPIAQT 44
Db 197 RPIAQT 203

RESULT 12
Q92L02      PRELIMINARY;      PRT;      257 AA.
ID Q92L02
AC Q92L02;
DT 01-DEC-2001 (TREMBlrel. 19, Created)
DT 01-DEC-2001 (TREMBlrel. 19, Last sequence update)
DT 01-MAR-2002 (TREMBlrel. 20, Last annotation update)
DE Putative oxidoreductase protein.
DR EMBL; AL591793; CAC47875.1; -.
DR InterPro: IPR002198; ADH_short.
DR Pfam; PF00106; adh_short_1.
DR PROSITE; PS00061; ADH_SHORT; UNKNOWN_1.
KM Complete proteome.
SQ SEQUENCE 257 AA; 26549 MW; C32CD2629A1BAF6E CRC64;

Query Match      6.6%; Score 7; DB 16; Length 257;
Best Local Similarity 100.0%; Pred. No. 38;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 45 SGPDIAI 51
Db 29 SGPDIAI 35

RESULT 13
Q8Y1KO      PRELIMINARY;      PRT;      300 AA.
ID Q8Y1KO
```

AC QBYIK0; 01-MAR-2002 (Tremblrel. 20, Created)  
DT 01-MAR-2002 (Tremblrel. 20, Last sequence update)  
DT 01-JUN-2002 (Tremblrel. 21, Last annotation update)  
DE Probable transmembrane protein.  
GN R5C0690 OR R501595.  
OS Ralstonia solanacearum (Pseudomonas solanacearum).  
OC Bacteria; Proteobacteria; beta subdivision; Ralstonia group;  
OC Ralstonia.  
OX NCBI\_TaxID=305;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=GM1100;  
RX MEDLINE=21681879; PubMed=11823852;  
RA Salanoubat M., Genin S., Artiguenave F., Guzy J., Mangenot S.,  
Ariat M., Billault A., Brottier P., Camus J.C., Gattolico L.,  
Chandler M., Choisme N., Claudel-Renard C., Cunnac S., Demange N.,  
Gaapin C., Lavie W., Moisan A., Robert C., Saurin W., Schlex T.,  
RA Siglier P., Thebaud P., Whalen M., Wincker P., Levy M.,  
Weissenbach J., Boucher C.A.;  
RT "Genome sequence of the plant pathogen Ralstonia solanacearum";  
RL Nature 415:497-502(2002)  
DR InterPro; IPR003856; W2Z.  
DR Pfam; PF02706; w2z; 1.  
KW Complete proteome.  
SQ SEQUENCE 300 AA; 32607 MW; PD9BLC0A9EA65B51 CRC64;

Query Match 6.6%; Score 7; DB 16; Length 300;  
Best Local Similarity 100.0%; Pred. No. 43;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 50 AIVMLAQ 56  
Db 16 AIVMLAQ 22

RESULT 14  
098H59 PRELIMINARY; PRT; 316 AA.  
AC 098H59;  
DT 01-OCT-2001 (Tremblrel. 18, Created)  
DT 01-OCT-2001 (Tremblrel. 18, Last sequence update)  
DT 01-MAR-2002 (Tremblrel. 20, Last annotation update)  
DE Hypothetical protein m1r3021.  
GN M1R3021.  
OS Rhizobium loti (Mesorhizobium loti).  
OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;  
OC Phyllobacteriaceae; Mesorhizobium.  
OX NCBI\_TaxID=381;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=MAFF303099;  
RX MEDLINE=21082930; PubMed=11214968;  
RA Kaneko T., Nakamura Y., Sato S., Asamizu E., Kato T., Sasamoto S.,  
Watanabe A., Idegawa K., Ishikawa A., Kawashima K., Kimura T.,  
RA Kishida Y., Kiyokawa C., Kohara M., Matsumoto M., Matsuno A.,  
Mochizuki Y., Nakayama S., Nakazaki N., Shimpoto S., Sugimoto M.,  
Takeuchi C., Yamada M., Tabata S.;  
RT "Complete genome structure of the nitrogen-fixing symbiotic bacterium  
Mesorhizobium loti";  
RL DNA Res. 7:331-338(2000).  
DR EMBL; AP003001; BAB50007.1; -.  
DR InterPro; IPR001107; Band 7.  
DR InterPro; IPR001972; Stomatol.  
DR Pfam; PF01145; Band 7; 1.  
DR PRINTS; PR00721; STOMATIN.  
DR SMART; SM00244; PHB; 1.  
KW Hypothetical protein; Complete proteome.  
SQ SEQUENCE 316 AA; 34513 MW; C55FDP289933692F CRC64;

Query Match 6.6%; Score 7; DB 16; Length 316;  
Best Local Similarity 100.0%; Pred. No. 45;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Qy 45 SGFDIAI 51  
Db 4 SGFDIAI 10

RESULT 15  
08RFR1 PRELIMINARY; PRT; 336 AA.  
AC 08RFR1;  
DT 01-JUN-2002 (Tremblrel. 21, Created)  
DT 01-JUN-2002 (Tremblrel. 21, Last sequence update)  
DT 01-JUN-2002 (Tremblrel. 21, Last annotation update)  
DE Glucosamine--fructose-6-phosphate aminotransferase (EC 2.6.1.16).  
GN FN0627.  
OS Fusobacterium nucleatum (subsp. nucleatum).  
OC Bacteria; Fusobacteriia; Fusobacterium.  
OX NCBI\_TaxID=76856;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=ATCC 25586;  
RX MEDLINE=21866394; PubMed=11889109;  
RA Kapatchay V., Anderson I., Ivanova N., Reznik G., Los T., Lykidis A.,  
Bhatnagar A., Barton A., Gardner W., Grechkin G., Zhu L.,  
RA Vasileva O., Chu L., Kogan Y., Chaga O., Goldsman E., Bernal A.,  
RA Larsen N., D'Souza M., Matunas T., Pusch G., Haselkorn R.,  
RA Fongstein M., Kyrides N., Overbeek R.;  
RT "Genome sequence and analysis of the oral bacterium Fusobacterium  
nucleatum strain ATCC 25586";  
RL J. Bacteriol. 184:2005-2018(2002).  
DR EMBL; AE010574; AL94823.1; -.  
KW Transferase; Aminotransferase; Complete proteome.  
SQ SEQUENCE 336 AA; 39110 MW; 6655E7A3BB82A221 CRC64;

Query Match 6.6%; Score 7; DB 16; Length 336;  
Best Local Similarity 100.0%; Pred. No. 48;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 63 GIRVISTL 69  
Db 110 GIRVISTL 116

RESULT 16  
P79783 PRELIMINARY; PRT; 342 AA.  
AC P79783;  
DT 01-MAY-1997 (Tremblrel. 03, Created)  
DT 01-MAY-1997 (Tremblrel. 03, Last sequence update)  
DT 01-MAR-2002 (Tremblrel. 20, Last annotation update)  
DE CMP-NeuAc:GM3 sialyltransferase.  
GN SIAL-T2.  
OS Gallus gallus (Chicken).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Archaeoptera; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;  
OC Gallus.  
OX NCBI\_TaxID=9031;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=96063819; PubMed=9402227;  
RA Danilov J.L., Rosales Filtz V., Kunda P., Nishi T., Maccioni H.J.;  
RT "Cloning, characterization and developmental expression of alpha2,8  
sialyltransferase (GD3 synthase, ST8Sia I) gene in chick brain and  
retina";  
RL Int. J. Dev. Neurosci. 15:767-776(1997).  
DR EMBL; U73176; AAC2888.1; -.  
DR InterPro; IPR001675; GT 29.  
DR Pfam; PF00777; Glyco transf 29; 1.  
KW Glycosyltransferase; Transferase.  
SQ SEQUENCE 342 AA; 38863 MW; 7B6FCD73C85AB6D8 CRC64;

Query Match 6.6%; Score 7; DB 13; Length 342;

Best Local Similarity 100.0%; Pred. No. 48;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 100 NGGILKK 106  
DB 129 NGGILKK 135

## RESULT 17

097123 PRELIMINARY; PRT; 359 AA.

AC 097123;  
DT 01-DEC-2001 (TREMBlrel. 19, Created)  
DT 01-DEC-2001 (TREMBlrel. 19, Last sequence update)  
DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)  
DE Hypothetical protein S71234.  
GN S71234.  
OS Sulfolobus tokodaii.  
OC Archaea; Crenarchaeota; Thermoprotei; Sulfolobales; Sulfolobaceae;  
OC Sulfolobus.  
OX NCBI\_TaxID=11955;  
RN [1]

RP SEQUENCE FROM N.A.  
RC STRAIN=JCM 10545 / 7;  
RX PubMed=11572479;

RA Kawabayashi Y., Hino Y., Horikawa H., Jin-no K., Takahashi M.,  
RA Sekine M., Baba S.-I., Anka A., Kougi H., Hosoyama A., Fukui S.,  
RA Nagai Y., Nishijima K., Otsuka R., Nakazawa H., Takamiya M., Kato Y.,  
RA Yoshizawa T., Tanaka T., Kudon Y., Yamazaki J., Kushida N., Oguchi A.,  
RA Aoki K.-I., Masuda S., Yanagii M., Nishimura M., Yamagishi A.,  
RA Oshima T., Kikuchi H.;  
RT "Complete genome sequence of an aerobic thermophilic  
RT Crenarchaeon, Sulfolobus tokodaii strain7.",  
RL DNA Res. 8:123-140(2001).  
DR EMBL; AP000985; BAB6276.1; -  
DR InterPro; IPR002587; Inos-1-P\_synth.  
DR InterPro; IPR00504; RNA\_rec\_mot.  
DR Pfam; PF01658; Inos-1-P\_synth; 1.  
DR PROSITE; PS00030; RRM\_RNP\_1; UNKNOWN\_1.  
KW Hypothetical protein; Complete proteome.  
SQ SEQUENCE 359 AA; 39615 MW; ED28B8AE10B375A4 CRC64;

Query Match 6.6%; Score 7; DB 17; Length 359;  
Best Local Similarity 100.0%; Pred. No. 50;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 20 GVVAVPG 26  
DB 85 GVVAVPG 91

RESULT 18  
052988 PRELIMINARY; PRT; 382 AA.

AC 052988;  
DT 01-JUN-1998 (TREMBlrel. 06, Created)  
DT 01-JUN-1998 (TREMBlrel. 06, Last sequence update)  
DT 01-MAR-2002 (TREMBlrel. 20, Last annotation update)  
DE ES-beta-lactamase precursor.  
GN BLA-FOX-2.  
OS Escherichia coli.  
OC Plasmid pMV-7.  
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;  
OC Escherichia.  
OX NCBI\_TaxID=562;  
RN [1]

RP SEQUENCE FROM N.A.  
RC STRAIN=BW U2206;  
RX MEDLINE=97447673; PubMed=9303413;

RA Bauerfeind A., Wagner S., Jungwirth R., Schneider J., Meyer D.;  
RT "A novel class C beta-lactamase (FOX-2) in Escherichia coli  
conferring resistance to cephamycins.";  
RL Antimicrob. Agents Chemother. 41:2041-2046(1997).

DR EMBL; Y10282; CAAT325.1; -.

DR HSSP; P05364; 2BLT.

DR InterPro; IPR001466; Beta\_lactamase.

DR InterPro; IPR001586; Beta\_lactamase\_C.

DR Pfam; PF00144; beta-lactamase; 1.

DR PROSITE; PS00336; BETA\_LACTAMASE\_C; 1.

KW Plasmid; Signal.

FT SIGNAL 1 23 POTENTIAL.

FT CHAIN 24 382 POTENTIAL.

SQ SEQUENCE 382 AA; 40935 MW; 13E3B3804035B885 CRC64;

Query Match 6.6%; Score 7; DB 2; Length 382;  
Best Local Similarity 100.0%; Pred. No. 53;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

## RESULT 19

091387 PRELIMINARY; PRT; 382 AA.

AC 091387;  
DT 01-OCT-2000 (TREMBlrel. 15, Created)  
DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)  
DT 01-MAR-2002 (TREMBlrel. 20, Last annotation update)  
DE Beta-lactamase class C.  
GN FOX-4.  
OS Escherichia coli.  
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;  
OC Escherichia.  
OX NCBI\_TaxID=562;  
RN [1]

RP SEQUENCE FROM N.A.  
RC STRAIN=GCE;  
RX MEDLINE=20408594; PubMed=10952615;  
RA Bou G., Oliver A., Ofeda M., Monzon C., Martinez-Beltran J.;  
RT "Molecular characterization of FOX-4, a new AmpC-type plasmid-mediated  
beta-lactamase from an Escherichia coli strain isolated in Spain.";  
RL Antimicrob. Agents Chemother. 44:2549-2553(2000).  
DR EMBL; AJ277535; CAB89086.1; -.

DR HSSP; P05364; 2BLT.  
DR InterPro; IPR001466; Beta\_lactamase.  
DR InterPro; IPR001586; Beta\_lactamase\_C.  
DR Pfam; PF00144; beta-lactamase; 1.  
DR PROSITE; PS00336; BETA\_LACTAMASE\_C; 1.

KW Plasmid.  
SQ SEQUENCE 382 AA; 40960 MW; 9476CE4C4A347A37 CRC64;

Query Match 6.6%; Score 7; DB 2; Length 382;  
Best Local Similarity 100.0%; Pred. No. 53;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 49 IAIWLA 55  
DB 354 IAIWLA 360

## RESULT 20

048437 PRELIMINARY; PRT; 382 AA.

AC 048437;  
DT 01-NOV-1996 (TREMBlrel. 01, Created)  
DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)  
DT 01-MAR-2002 (TREMBlrel. 20, Last annotation update)  
DE Beta-lactamase, class C.

GN FOX.  
OS Klebsiella pneumoniae.  
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;  
OC Klebsiella.  
OX NCBI\_TaxID=573;

[1]  
RN SEQUENCE FROM N.A.  
RP STRAIN=BA32;  
RC MEDLINE=95110083; PubMed=7811034;  
RA Gonzalez-Leiza M., Perez-Diaz J.C., Ayala J., Casellas J.M.,  
RT Martinez-Beltran J., Bush K., Baquero F.;  
RT "Gene sequence and biochemical characterization of FOX-1 from  
RT Klebsiella pneumoniae, a new AmpC-type plasmid-mediated beta-lactamase  
RT with two molecular variants.";  
RL Antimicrob. Agents Chemother. 38:2150-2157(1994).  
DR EMBL; X77455; CAA54602.1; -.  
DR HSSP; P05364; 2BLT.  
DR InterPro; IPR001466; Beta\_lactamase.  
DR InterPro; IPR001586; Beta\_lactamase\_C.  
DR Pfam; PF00144; beta\_lactamase\_1.  
DR PROSITE; PS00336; BETA\_LACTAMASE\_C; 1.  
SQ SEQUENCE 382 AA; 41191 MW; C2516320E9357B9F CRC64;  
  
Query Match 6.6%; Score 7; DB 2; Length 382;  
Best Local Similarity 100.0%; Pred. No. 53;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
OY 49 IAIWMLA 55  
DB 354 IAIWMLA 360  
  
RESULT 21  
P94856 PRELIMINARY; PRT; 382 AA.  
AC P94856;  
DT 01-MAY-1997 (TREMBLrel. 03, Created)  
DT 01-MAY-1997 (TREMBLrel. 03, Last sequence update)  
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)  
DE Beta-lactamase class C.  
GN BLA FOX-3.  
OS Klebsiella oxytoca.  
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;  
OC Klebsiella.  
OX NCBI\_Taxid=571;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=1731;  
RX MEDLINE=98139051; PubMed=9527810;  
RA Marchese A., Arlet G., Schito G.C., Lagrange P.H., Philippou A.;  
RT "Characterization of FOX-3, an AmpC-type plasmid-mediated beta-  
RT lactamase from Italian Klebsiella oxytoca and Klebsiella pneumoniae  
RT isolates.";  
RL Antimicrob. Agents Chemother. 42:464-467(1998).  
DR EMBL; Y11068; CAA71947.1; -.  
DR HSSP; P05364; 2BLT.  
DR InterPro; IPR001466; Beta\_lactamase.  
DR InterPro; IPR001586; Beta\_lactamase\_C.  
DR Pfam; PF00144; beta\_lactamase\_1.  
DR PROSITE; PS00336; BETA\_LACTAMASE\_C; 1.  
SQ SEQUENCE 382 AA; 40993 MW; E38BD822C2F41EF CRC64;  
  
Query Match 6.6%; Score 7; DB 2; Length 382;  
Best Local Similarity 100.0%; Pred. No. 53;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
OY 49 IAIWMLA 55  
DB 354 IAIWMLA 360  
  
RESULT 22  
O9F4K3 PRELIMINARY; PRT; 382 AA.  
AC O9F4K3;  
DT 01-MAR-2001 (TREMBLrel. 16, Created)  
DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)  
DT 01-MAR-2002 (TREMBLrel. 20, Last annotation update)

DE Beta-lactamase FOX-5.  
GN FOX-6.  
OS Klebsiella pneumoniae.  
OC Plasmid p5064.  
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;  
OC Klebsiella.  
OX NCBI\_Taxid=573;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC PLASMID=p5064;  
RA Queenan A.-M., Bush K., Jenkins S.G.;  
RT "A clavulanate-resistant K. pneumoniae clinical isolate encoding  
RT plasmid-mediated TEM-2 like and FOX-5 beta-lactamases.";  
RL Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.  
RN [2]  
RP SEQUENCE FROM N.A.  
RA Herrera V.L., Hanson N.D.;  
RT "Fox6, A Novel Plasmid-mediated AmpC Beta-lactamase.";  
RL Submitted (MAY-2001) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AY007369; AAG12974.1; -.  
DR EMBL; AY034848; AAK70221.1; -.  
DR HSSP; P05364; 2BLT.  
DR InterPro; IPR001466; Beta\_lactamase.  
DR InterPro; IPR001586; Beta\_lactamase\_C.  
DR Pfam; PF00144; beta\_lactamase\_1.  
DR PROSITE; PS00336; BETA\_LACTAMASE\_C; 1.  
KW Plasmid.  
SQ SEQUENCE 382 AA; 40882 MW; BBF905F30DEF2B8C CRC64;  
  
Query Match 6.6%; Score 7; DB 2; Length 382;  
Best Local Similarity 100.0%; Pred. No. 53;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
OY 49 IAIWMLA 55  
DB 354 IAIWMLA 360  
  
RESULT 23  
O98K89 PRELIMINARY; PRT; 406 AA.  
AC O98K89;  
DT 01-OCT-2001 (TREMBLrel. 18, Created)  
DT 01-OCT-2001 (TREMBLrel. 18, Last sequence update)  
DT 01-MAR-2002 (TREMBLrel. 20, Last annotation update)  
DE Ornithine decarboxylase.  
GN MLI584.  
OS Rhizobium loti (Mesorhizobium loti)  
OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;  
OC Phyllobacteriaceae; Mesorhizobium.  
OX NCBI\_Taxid=381;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=MAFF303099;  
RX MEDLINE=21082930; PubMed=11214968;  
RA Kaneko T., Nakamura Y., Sato S., Asamizu E., Kato T., Sasamoto S.,  
RA Watanabe A., Idegawa K., Ishikawa A., Kawashina K., Kimura T.,  
RA Kishida Y., Kiyokawa C., Kohara M., Matsumoto M., Matsuno A.,  
RA Mochizuki Y., Nakayama S., Nakazaki N., Shimo S., Sugimoto M.,  
RA Takeuchi C., Yamada M., Tabata S.;  
RT "Complete genome structure of the nitrogen-fixing symbiotic bacterium  
RT Mesorhizobium loti.";  
RL DNA Res. 7:331-338(2000).  
DR EMBL; AP002997; BAB48925.1; -.  
DR InterPro; IPR000183; Decarboxylase.  
DR Pfam; PF02784; Orn\_Arg\_dec\_N; 1.  
DR Pfam; PF00278; Orn\_DAP\_Arg\_dec; 1.  
DR PRINTS; PR01179; ODADCRBLXASE.  
KW Complete proteome.  
SQ SEQUENCE 406 AA; 44866 MW; CG337BDE5E3CD6B9C CRC64;  
  
Query Match 6.6%; Score 7; DB 16; Length 406;  
Best Local Similarity 100.0%; Pred. No. 56;



Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 63 GIRVISTL 69  
|||||  
Db 107 GIRVISTL 113

## RESULT 24

Q9FGZ7 PRELIMINARY; PRT; 461 AA.  
AC Q9FGZ7;  
DT 01-MAR-2001 (TREMBlrel. 16, Created)  
DT 01-MAR-2001 (TREMBlrel. 16, Last sequence update)  
DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)  
DE Genomic DNA, chromosome 5, TAC clone:K5J14.  
OS Arabidopsis thaliana (Mouse-ear cress).  
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;  
OC eurosid 11; Brassicales; Brassicaceae; Arabidopsie.  
OC NCBI\_TaxID=3702;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=COLUMBIA;  
RX MEDLINE=20181125; PubMed=10718197;  
RA Sato S., Nakamura Y., Kaneko T., Katoh T., Asamizu E., Kotani H.,  
RA Tabata S.;  
RT "Structural analysis of Arabidopsis thaliana chromosome 5. X. Sequence  
RT features of the regions of 3,076,755 bp covered by sixty P1 and TAC  
RT clones.";  
RL DNA Res. 7:31-63 (2000).  
DR EMBL; AB023032; BAB10205.1;  
DR EMBL; AB016888; BAB10205.1; JOINED.  
DR InterPro: IPR000834; Zn\_carpodect.  
DR Pfam: PF00246; Zn\_carpodect.1.  
SQ SEQUENCE 461 AA; 52612 MW; DE9BE2937A06734C CRC64;

Query Match  
Best Local Similarity 100.0%; Score 7; DB 10; Length 461;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 99 KNGGILK 105  
|||||  
Db 166 KNGGILK 172

## RESULT 25

Q9Y1X5 PRELIMINARY; PRT; 478 AA.  
AC Q9Y1X5;  
DT 01-NOV-1999 (TREMBlrel. 12, Created)  
DT 01-NOV-1999 (TREMBlrel. 12, Last sequence update)  
DT 01-MAR-2002 (TREMBlrel. 20, Last annotation update)  
DE SPTPR2B (fragment).  
OS Ephydratia fluviatilis.  
OC Eukaryota; Metazoa; Porifera; Demospongiae; Ceractinomorpha;  
OC Haploclerida; Spongillidae; Ephydratia.  
OC NCBI\_TaxID=31330;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=99246376; PubMed=10229569;  
RA Ono K., Suga H., Iwabe H., Kuma K., Miyata T.;  
RT "Multiple protein tyrosine phosphatases in sponges and explosive gene  
RT duplication in the early evolution of animals before the parazoan-  
RT eumetazoan split.";  
RT U. Mol. Evol. 48:654-662 (1999).  
DR EMBL; AB019126; BAA82559.1;  
DR HSSP; P18052; 1YFO.  
DR InterPro: IPR000387; TYR\_phosphatase.  
DR InterPro: IPR000242; TYR\_PP.  
DR Pfam; PF00102; Y\_phosphatase.2.  
DR PRINTS; PR00700; PRTYPPHATASE.  
DR SMART; SM00194; PTPc; 2.  
DR PROSITE; PS00383; TYR\_PHOSPHATASE\_1; 2.

DR PROSITE; PSS0056; TYR\_PHOSPHATASE\_2; 2.  
DR PROSITE; PSS0055; TYR\_PHOSPHATASE\_PP; 2.  
KW Hydrolase.  
FT NON TER  
SQ SEQUENCE 478 AA; 55019 MW; EE30275F6664DB9A CRC64;

Query Match  
Best Local Similarity 100.0%; Score 7; DB 5; Length 478;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 50 AIWMLAQ 56  
|||||  
Db 300 AIWMLAQ 306

## RESULT 26

Q9URL7 PRELIMINARY; PRT; 623 AA.  
AC Q9URL7;  
DT 01-MAY-2000 (TREMBlrel. 13, Created)  
DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)  
DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)  
DE Peptide transport protein.  
GN PTR2.  
OS Candida albicans (Yeast).  
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
OC Saccharomycetales; mitosporic Saccharomycetales; Candida.  
OC NCBI\_TaxID=5476;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=1161;  
RA Murphy L., Harris D.;  
RL Submitted (NOV-1998) to the EMBL/GenBank/DBJ databases.  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=1161;  
RA Barrell B.G., Rajandream M.A.;  
RL Submitted (NOV-1998) to the EMBL/GenBank/DBJ databases.  
RN [3]  
RP SEQUENCE FROM N.A.  
RC STRAIN=1161;  
RX MEDLINE=97435544; PubMed=9290243;  
RA Tait E., Simon M.C., King S., Brown A.J., Gow N.A.R., Shaw D.J.;  
RT "A Candida albicans Genome Project: Cosmid Contigs, Physical Mapping,  
RT and Gene Isolation.";  
RL Fungal Genet. Biol. 21:308-314 (1997).  
DR EMBL; AL033503; CAA22021.1;  
DR InterPro: IPR000109; PTR2.  
DR Pfam; PF00854; PTR2.1.  
DR PROSITE; PS01023; PTR2\_1; 1.  
DR PROSITE; PS01023; PTR2\_2; UNKNOWN.1.  
SQ SEQUENCE 623 AA; 69943 MW; 8543A3B1F7E7363E CRC64;

Query Match  
Best Local Similarity 100.0%; Score 7; DB 3; Length 623;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 50 AIWMLAQ 56  
|||||  
Db 572 AIWMLAQ 578

## RESULT 27

Q9XE83 PRELIMINARY; PRT; 657 AA.  
AC Q9XE83;  
DT 01-NOV-1999 (TREMBlrel. 12, Created)  
DT 01-NOV-1999 (TREMBlrel. 12, Last sequence update)  
DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)  
DE Serine carboxypeptidase-like protein.  
OS Sorghum bicolor (Sorghum) (Sorghum vulgare).  
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC clade;

```
OC Panicoidae; Andropogoneae; Sorghum.
OX NCBI_TaxID=4558;
RN [1]
RP SEQUENCE FROM N.A.
RA Ilaca V., Lou A., Messing J.W.;
RT "Microsytenuy analysis of 22-kda zein cluster in maize and sorghum.";
RL Submitted (MAR-1999) to the EMBL/Genbank/DBJ databases.
DR EMBL; AF061282; AAD2151.1; -.
DR HSSP; P08819; 1MHT.
DR MEROPS; S10.005; -.
DR InterPro; IPR001563; Serine_carpept.
DR InterPro; IPR000379; Ser_ester_site.
DR Pfam; PF00450; serine_carpept; 1.
DR ProDom; PD001189; Serine_carpept; 1.
DR PROSITE; PS00131; CARBOXYPEPT_SER_SER; UNKNOWN_1.
KM Carboxypeptidase.
SQ SEQUENCE 657 AA; 74035 MW; DD120F4BAE703155 CRC64;

Query Match 6.6%; Score 7; DB 10; Length 657;
Best Local Similarity 100.0%; Pred. No. 83;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 76 PPGTGV 82
DB 216 PPGTGV 222

RESULT 28
OQ9A6Y0 PRELIMINARY; PRT; 686 AA.
ID OQ9A6Y0;
AC OQ9A6Y0;
DT 01-JUN-2001 (TREMBLrel. 17, Created)
DT 01-JUN-2001 (TREMBLrel. 17, Last sequence update)
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE MAH dehydrogenase I, G subunit.
GN CC1946.
OS Caulobacter crescentus.
OC Bacteria; Proteobacteria; alpha subdivision; Caulobacter group;
OC Caulobacter.
OX NCBI_TaxID=155892;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 19089 / CB15;
RX MEDLINE=21173698; PubMed=11259647;
RA Nieman W.C., Feldblum T.V., Laub M.T., Paulsen I.T., Nelson K.E.,
RA Eisen J., Heidelberg J.F., Alley M.R.K., Ohta N., Maddock J.R.,
RA Potocka I., Nelson W.C., Newton A., Stephens C., Phadke N.D., Ely B.,
RA DeJoy R.T., Dodson R.J., Durkin A.S., Gwinn M.L., Haft D.H.,
RA Kolonay J.F., Smit J., Craven M.B., Khouli H., Shetty J., Berry K.,
RA Utechtback T., Tran K., Wolf A., Vamathevan J., Ermolaeva M., White O.,
RA Salzberg S.L., Venter J.C., Shapiro L., Fraser C.M.;
RT "Complete genome sequence of Caulobacter crescentus.";
RL Proc. Natl. Acad. Sci. U.S.A. 98:4136-4141(2001).
DR EMBL; AE005868; AAK23921.1; -.
DR TIGR; CC1946; -.
DR InterPro; IPR000283; Complex1_75K.
DR InterPro; IPR001041; Ferridoxin.
DR InterPro; IPR001467; Prok_Mboxred.
DR Pfam; PF00111; fer2; 1.
DR Pfam; PF00384; molycopdoperin; 1.
DR PROSITE; PS00641; COMPLEX1_75K_1; 1.
DR PROSITE; PS00642; COMPLEX1_75K_2; 1.
DR PROSITE; PS00643; COMPLEX1_75K_3; 1.
KW Iron-sulfur; Complete proteome.
SQ SEQUENCE 686 AA; 73163 MW; 3B4955F60FB55F70 CRC64;

Query Match 6.6%; Score 7; DB 16; Length 686;
Best Local Similarity 100.0%; Pred. No. 86;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 85 VGYGRD 91
DB 115 VGYGRD 121
```

```
RESULT 29
OQ8ZYH4 PRELIMINARY; PRT; 794 AA.
ID OQ8ZYH4;
AC OQ8ZYH4;
DT 01-MAR-2002 (TREMBLrel. 20, Created)
DT 01-MAR-2002 (TREMBLrel. 20, Last sequence update)
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE Hypothetical protein PAE2280.
GN PAE2280.
OS Pyrobaculum aerophilum.
OC Archaea; Crenarchaeota; Thermoprotei; Thermoproteales;
OC Thermoproteaceae; Pyrobaculum.
OX NCBI_TaxID=13773;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=IM2 / ATCC 51768 / DSM 7523;
RX PubMed=11792869;
RA Fitz-Gibbon S.T., Ladner H., Kim U.-J., Stetter K.O., Simon M.I.,
RA Miller J.H.;
RT "Genome sequence of the hyperthermophilic crenarchaeon Pyrobaculum
aerophilum.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:984-989(2002).
DR EMBL; AE009867; AAL64082.1; -.
DR InterPro; IPR002106; AAcRNA_ligaseII.
DR InterPro; IPR003439; ABC_transport.
DR InterPro; IPR000719; Euk_pkinase.
DR InterPro; IPR003395; SMC_N.
DR Pfam; PF02463; SMC_N; 1.
DR PROSITE; PS00339; AA_TENA_LIGASE_II_2; UNKNOWN_1.
DR PROSITE; PS00107; PROTEIN_KINASE_ATP; UNKNOWN_1.
KM Hypothetical protein; Complete proteome.
SQ SEQUENCE 794 AA; 91123 MW; 530347055A96034C CRC64;

Query Match 6.6%; Score 7; DB 17; Length 794;
Best Local Similarity 100.0%; Pred. No. 98;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 77 LTRNGD 83
DB 77 LTRNGD 83
```

```
RESULT 30
OQW1S4 PRELIMINARY; PRT; 838 AA.
ID OQW1S4;
AC OQW1S4;
DT 01-MAY-2000 (TREMBLrel. 13, Created)
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
DT 01-JUN-2000 (TREMBLrel. 14, Last annotation update)
DE CG9803 protein.
GN CG9803.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BERKELEY;
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazey R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abril J.F., Aghayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA Ballew R.M., Baau J., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Bernos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borkova D., Botchan M.R., Bouck U., Brokstein P., Brotler P.,
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
```

RA Cherry J.M., Cavley S., Dahlke C., Davenport L.B., Davies P.,  
 RA de Pablo B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,  
 RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,  
 RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,  
 RA Foeller C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glasser K.,  
 RA Glodet A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,  
 RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,  
 RA Hostin D., Houston K.A., Howland T.U., Wei M.-H., Ibeagwa C.,  
 RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,  
 RA Kammel B.E., Kodira C.D., Kraft C., Kravitz S., Kalp D., Lai Z.,  
 RA Lau X., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,  
 RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,  
 RA Mekulov G., Mishina N.V., Mobarry C., Morris J., Moshrefi A.,  
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,  
 RA Nelson D.R., Nelson K.A., Nixon K., Nuskern D.R., Pacle J.M.,  
 RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,  
 RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,  
 RA Shue B.C., Siden-Kimios I., Simpson M., Skupski M.P., Smith T.,  
 RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,  
 RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,  
 RA Wang Z.-Y., Wasserman D.A., Weinstein G.M., Weisenbach J.,  
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,  
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao C., Zheng L.,  
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,  
 RA Gdbbs R.A., Myers E.W., Rubin G.M., Venter J.C.;  
 RT "The genome sequence of *Drosophila melanogaster*.";  
 RL Science 287:2185-2195(2000).  
 DR EMBL: AE003460; AAF6979.1; -  
 DR FLYBASE: FBgn0034843; CG9803.  
 SQ SEQUENCE 838 AA; 95949 MW; 716806DCB7CEBEE CRC64;

Query Match 6.6%; Score 7; DB 5; Length 838;  
 Best Local Similarity 100.0%; Pred. No. 1e+02;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 72 PSDIPP 78  
 |||||  
 Db 300 PSDIPP 306

RESULT 31  
 Q9NT68 PRELIMINARY; PRT; 849 AA.  
 ID Q9NT68  
 AC Q9NT68;  
 DT 01-OCT-2000 (TrEMBLrel. 15, Created)  
 DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)  
 DT 01-OCT-2000 (TrEMBLrel. 15, Last annotation update)  
 DE Hypothetical 97.1 kDa protein (Fragment).  
 GN DKFZP761F171.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
 OX NCB1\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=AVYGDAL;  
 RA Ostermeier B., Ostermeier B., Mewes H.W., Gassenhuber J., Wiemann S.;  
 RL Submitted (JAN-2000) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: AL137500; CAB70774.1; -  
 KM Hypothetical protein.  
 FT NON\_TER 1  
 SQ SEQUENCE 849 AA; 97142 MW; F7EF309A9067A7B8 CRC64;

Query Match 6.6%; Score 7; DB 4; Length 849;  
 Best Local Similarity 100.0%; Pred. No. 1e+02;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 57 MVNLOSG 63  
 |||||  
 Db 126 MVNLOSG 132

RESULT 32

Q41058 PRELIMINARY; PRT; 922 AA.  
 ID Q41058  
 AC Q41058;  
 DT 01-NOV-1996 (TrEMBLrel. 01, Created)  
 DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)  
 DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)  
 DE Starch branching enzyme I precursor.  
 GN SBEI.  
 OS Pisum sativum (Garden pea).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;  
 OC eucosids I; Fabales; Fabaceae; Papilionoideae; Viciae; Pisum.  
 OX NCB1\_TaxID=3888;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=95201826; PubMed=7894509;  
 RA Burton R.A., Bewley J.D., Smith A.M., Bhattacharya M.K., Tatge H.,  
 RA Ring S., Bull V., Hamilton W.D.O., Martin C.;  
 RT "Starch branching enzymes belonging to distinct enzyme families are  
 differentially expressed during pea embryo development.";  
 RL Plant J. 7:3-15(1995).  
 DR EMBL: X80009; CA56319.1; -  
 DR InterPro: IPR000461; Alpha\_amyase.  
 DR InterPro: IPR004193; Isoamylase\_N.  
 DR Pfam: PF00128; alpha-amyase; 1.  
 DR Pfam: PF02922; isoamylase\_N; 1.  
 KM Transit peptide.  
 FT TRANSIT 1  
 FT CHAIN 49  
 SQ SEQUENCE 922 AA; 105227 MW; 4EDBF9374C6385C8 CRC64;

Query Match 6.6%; Score 7; DB 10; Length 922;  
 Best Local Similarity 100.0%; Pred. No. 1e+02;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 75 IPPPGTG 81  
 |||||  
 Db 159 IPPPGTG 165

RESULT 33  
 Q9JLCO PRELIMINARY; PRT; 930 AA.  
 ID Q9JLCO  
 AC Q9JLCO;  
 DT 01-OCT-2000 (TrEMBLrel. 15, Created)  
 DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)  
 DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)  
 DE OD22 (Fragment).  
 GN OD22.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OX NCB1\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Ben-Zur T., Feige E., Motro B., Wides R.;  
 RT "The mammalian Odz gene family: Homologs of a *Drosophila* pair rule  
 gene with expression implying distinct yet overlapping developmental  
 roles.";  
 RL Dev. Biol. 117:107-120(2000).  
 DR EMBL: AF195419; AAF28317.1; -  
 DR MGD: MGI:1345164; Odz2.  
 FT NON\_TER 1  
 SQ SEQUENCE 930 AA; 106506 MW; 9B879795182369DD CRC64;

Query Match 6.6%; Score 7; DB 11; Length 930;  
 Best Local Similarity 100.0%; Pred. No. 1e+02;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 57 MVNLOSG 63  
 |||||  
 Db 207 MVNLOSG 213

```

RESULT 34
Q63624      PRELIMINARY;      PRT; 1173 AA.
ID 063624
AC 063624
DT 01-NOV-1996 (TREMBLrel. 01, Created)
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE RA1.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=HIPPOCAMPUS;
RX MEDLINE=96293459; PubMed=8692929;
RA Yurjev A., Patutajan M., Litington Y., Joshi R.V., Gentile C.,
RA Gebara M., Corden J.L.;
RT "The C-terminal domain of the largest subunit of RNA polymerase II
RT interacts with a novel set of serine/arginine-rich proteins.";
RL Proc. Natl. Acad. Sci. U.S.A. 93:6975-6980(1996).
DR EMBL: U49056; AAC52657.1; -.
SQ SEQUENCE 1173 AA; 124775 MW; 2F5209A6435F6770 CRC64;

Query Match      6.6%; Score 7; DB 11; Length 1173;
Best Local Similarity 100.0%; Pred. No. 1.4e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 20 GVKVAPG 26
Db 751 GVKVAPG 757

RESULT 35
Q8T1G8      PRELIMINARY;      PRT; 1681 AA.
ID Q8T1G8
AC Q8T1G8;
DT 01-JUN-2002 (TREMBLrel. 21, Created)
DT 01-JUN-2002 (TREMBLrel. 21, Last sequence update)
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE UDP-glucose:glycoprotein glucosyltransferase.
OS Dictyostelium discoideum (Slime mold).
OC Eukaryota; Mycetozoa; Dictyostelida; Dictyostelium.
OX NCBI_TaxID=44689;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=AX4;
RA Gloeckner G., Eichinger L., Szafranski K., Pachebat J., Dear P.,
RA Lehman R., Baumgart C., Parra G., April J.F., Guigo R., Kumpf K.,
RA Tünggal B., Cox E., Quail M.A., Platzer M., Rosenthal A., Noegel A.A.;
RT "Sequence and Analysis of Chromosome 2 of Dictyostelium";
RL Submitted (MAR-2002) to the EMBL/Genbank/DBJ databases.
DR EMBL: AC116548; AAL93322.1; -.
KW TRANSFERASE.
SQ SEQUENCE 1681 AA; 192627 MW; 946F082CE2E58B18 CRC64;

Query Match      6.6%; Score 7; DB 5; Length 1681;
Best Local Similarity 100.0%; Pred. No. 1.8e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 44 LSGFDIA 50
Db 917 LSGFDIA 923

RESULT 36
Q9ULU2      PRELIMINARY;      PRT; 1737 AA.
ID Q9ULU2
AC Q9ULU2;
DT 01-MAY-2000 (TREMBLrel. 13, Created)
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)

```

```

DE KIAA1127 protein (Fragment).
GN KIAA1127.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=BRAIN;
RX MEDLINE=20039618; PubMed=10574461;
RA Hirosewa M., Nagase T., Ishikawa K., Kikuno R., Nomura N., Ohara O.;
RT "Characterization of cDNA clones selected by the Genemark analysis
RT from size-fractionated cDNA libraries from human brain.";
RL DNA Res. 6:329-336(1999).
DR EMBL: AB032953; BAA86441.1; -.
DR InterPro: IPR001258; NHL.
DR Pfam: PF01436; NHL; 6.
FT NON_TPR 1
SQ SEQUENCE 1737 AA; 195065 MW; CFF0AF0C8C5D6B32 CRC64;

Query Match      6.6%; Score 7; DB 4; Length 1737;
Best Local Similarity 100.0%; Pred. No. 1.9e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 57 MVNLQSG 63
Db 1014 MVNLQSG 1020

RESULT 37
Q9WTS5      PRELIMINARY;      PRT; 2764 AA.
ID Q9WTS5
AC Q9WTS5;
DT 01-NOV-1999 (TREMBLrel. 12, Created)
DT 01-NOV-1999 (TREMBLrel. 12, Last sequence update)
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE Ten-m2.
GN OD22 OR TEN-M2.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BALB/C; TISSUE=BRAIN;
RA Ohashi T., Zhou X., Peng K., Richter B., Moergelin M., Perez M.T.,
RA Su W., Chiquet-Ehrismann R., Rauch U., Faessler R.;
RT "Mouse Ten-m/Od2 is a new family of dimeric type II transmembrane
RT proteins expressed in many tissues.";
RL J. Cell Biol. 0:0-0(1999).
DR EMBL: AB025411; BAA77397.1; -.
DR HSSP: P35555; IEMN.
DR MED: MG1:1345184; Od22.
DR InterPro: IPR000561; EGF-like
DR InterPro: IPR002049; laminin_EGF.
DR InterPro: IPR001258; NHL.
DR Pfam: PF00008; EGF; 4.
DR Pfam: PF01436; NHL; 6.
DR PRINTS: PR00011; EGF1AMININ.
DR SMART: SM00181; EGF; 5.
DR PROSITE: PS00022; EGF_1; UNKNOWN_8.
DR PROSITE: PS01186; EGF_2; 7.
KW EGF-like domain; Glycoprotein.
SQ SEQUENCE 2764 AA; 306464 MW; 73BA3D916D0F0344 CRC64;

Query Match      6.6%; Score 7; DB 11; Length 2764;
Best Local Similarity 100.0%; Pred. No. 2.8e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 57 MVNLQSG 63
Db 2041 MVNLQSG 2047

```

```

RESULT 38
Q9RIK2      PRELIMINARY;      PRT; 2765 AA.
ID  Q9RIK2
AC  Q9RIK2;
DT  01-MAY-2000 (TrEMBLrel. 13, Created)
DT  01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT  01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE  Neurastin alpha.
OS  Rattus norvegicus (Rat).
OC  Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC  Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX  NCBI_TaxID=10116;
RN  [1]
RP  SEQUENCE FROM N.A.
RC  STRAIN=SPRAGUE DAWLEY; TISSUE=OLFACTORY BULB;
RX  MEDLINE=99350226; PubMed=10419693;
RA  Otsaki J.M., Friesstein S.;
RT  "Neurestin: putative transmembrane molecule implicated in neuronal
RT  development.";
RL  Dev. Biol. 212:165-181(1999).
RL  EMBL: AF086607; AAD47383.1; -.
DR  HSP; P00750; ITPG.
DR  InterPro; IPR000561; EGF-like.
DR  InterPro; IPR002049; Laminin_EGF.
DR  InterPro; IPR01258; NHL.
DR  Pfam; PF00008; EGF; 4.
DR  Pfam; PF01436; NHL; 6.
DR  PRINTS; PR00011; EGF/LAMININ.
DR  SMART; SM00181; EGF; 5.
DR  PROSITE; PS00022; EGF_1; UNKNOWN_8.
DR  PROSITE; PS01186; EGF_2; 7.
KM  EGF-like domain; Glycoprotein.
SQ  SEQUENCE 2765 AA; 306568 MW; 6748D70D5FFDE0E CRC64;

Query Match
Best Local Similarity 6.6%; Score 7; DB 11; Length 2765;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY  57 MVNLQSG 63
Db  2042 MVNLQSG 2048

RESULT 39
Q9DERS      PRELIMINARY;      PRT; 2802 AA.
ID  Q9DERS
AC  Q9DERS;
DT  01-MAR-2001 (TrEMBLrel. 16, Created)
DT  01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT  01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE  Teneurin-2 (Fragment).
GN  TEN2.
OS  Gallus gallus (Chicken).
OC  Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC  Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC  Gallus.
OX  NCBI_TaxID=9031;
RN  [1]
RP  SEQUENCE FROM N.A.
RX  MEDLINE=20581705; PubMed=11146505;
RA  Tucker R.P., Chiquet-Ehrismann R., Cheveron M., Martin D., Hall R.J.,
RA  Rubin B.P.;
RT  "Teneurin-2 is expressed in tissues that regulate limb and somite
RT  pattern formation and is induced in vitro and in situ by FGF8.";
RL  Dev. Dyn. 220:27-39(2001).
RL  EMBL: AJ279031; CAC09416.1; -.
DR  HSP; P00750; ITPG.
DR  InterPro; IPR000561; EGF-like.
DR  InterPro; IPR002049; Laminin_EGF.
DR  InterPro; IPR01258; NHL.
DR  Pfam; PF00008; EGF; 5.
DR  Pfam; PF01436; NHL; 6.

```

```

DR  PRINTS; PR00011; EGF/LAMININ.
DR  SMART; SM00181; EGF; 7.
DR  PROSITE; PS00022; EGF_1; UNKNOWN_7.
DR  PROSITE; PS01186; EGF_2; 7.
FT  NON_TER 1
FT  NON_TER 2802
SQ  SEQUENCE 2802 AA; 310745 MW; B1FBC2C84EDFA4B3 CRC64;

Query Match
Best Local Similarity 6.6%; Score 7; DB 13; Length 2802;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY  57 MVNLQSG 63
Db  2079 MVNLQSG 2085

RESULT 40
ID  O85558      PRELIMINARY;      PRT; 22 AA.
ID  O85558
AC  O85558;
DT  01-NOV-1998 (TrEMBLrel. 08, Created)
DT  01-NOV-1998 (TrEMBLrel. 08, Last sequence update)
DT  01-NOV-1998 (TrEMBLrel. 08, Last annotation update)
DE  Major outer membrane protein (Fragment).
GN  OMP-1.
OS  Chlamydia trachomatis.
OC  Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.
OX  NCBI_TaxID=813;
RN  [1]
RP  SEQUENCE FROM N.A.
RC  STRAIN=REF A2B;
RA  Hsieh Y.-H., Bobo L.D.;
RT  "Diversity of major outer membrane protein (omp-1) of Chlamydia
RT  trachomatis in trachoma endemic villages, Kongwa, Tanzania.";
RL  Submitted (JUN-1998) to the EMBL/GenBank/DBJ databases.
DR  EMBL: AF070356; AAC25327.1; -.
FT  NON_TER 1
FT  NON_TER 22
SQ  SEQUENCE 22 AA; 2290 MW; 8A2908AD7E7EB01E CRC64;

Query Match
Best Local Similarity 5.7%; Score 6; DB 2; Length 22;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY  45 SGFDIA 50
Db  6 SGFDIA 11

Search completed: April 9, 2003, 13:15:09
Job time : 132.098 secs

```

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.4 p5.4578  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 9, 2003, 12:59:52 ; Search time 25.1667 Seconds  
(without alignments)  
174.487 Million cell updates/sec

Title: US-10-020-441-2

Perfect score: 106  
Sequence: 1 VSFLLTNNGDQGIHQPSG.....YGRDDNDPDRKNGKILKK 106

Scoring table: OLIGO  
Gapop 60.0 , Gapext 60.0

Searched: 112892 seqs, 41476328 residues

Word size: 0

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 1000 summaries

Database: SwissProt\_40.\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	106	100.0	264	1	CERC_SCHMA
2	8	7.5	219	1	YE03_ARCFU
3	7	6.6	268	1	TRYP_STRGA
4	7	6.6	302	1	LDHB_RHIOA
5	7	6.6	356	1	S18A_HUMAN
6	7	6.6	420	1	MTD1_NEIMC
7	7	6.6	623	1	PTR2_CANAL
8	7	6.6	656	1	YK79_MYCTU
9	7	6.6	1121	1	MYT1_HUMAN
10	7	6.6	2300	1	CYAA_NEUCR
11	6	5.7	72	1	VXIS_BP434
12	6	5.7	72	1	VXIS_LAMBD
13	6	5.7	90	1	Y10K_MDV
14	6	5.7	97	1	Y123_ECOLI
15	6	5.7	102	1	CMGC_BACDH
16	6	5.7	112	1	YN48_ARCFU
17	6	5.7	116	1	SRI4_CAREL
18	6	5.7	122	1	YM63_ARCFU
19	6	5.7	144	1	RS15_HUMAN
20	6	5.7	149	1	Y881_VIBCH
21	6	5.7	153	1	MYG_AOTTR
22	6	5.7	153	1	MYG_BALAC
23	6	5.7	153	1	MYG_BALPH
24	6	5.7	153	1	MYG_BOVIN
25	6	5.7	153	1	MYG_CALJA
26	6	5.7	153	1	MYG_CANFA
27	6	5.7	153	1	MYG_CERAP
28	6	5.7	153	1	MYG_CEREL
29	6	5.7	153	1	MYG_ELEMA
30	6	5.7	153	1	MYG_BRIEU
31	6	5.7	153	1	MYG_ESCGI
32	6	5.7	153	1	MYG_GALCR
33	6	5.7	153	1	MYG_GORBE

34	6	5.7	153	1	MYG_HALGR
35	6	5.7	153	1	MYG_HORSE
36	6	5.7	153	1	MYG_HUMAN
37	6	5.7	153	1	MYG_HYLAG
38	6	5.7	153	1	MYG_TNIGE
39	6	5.7	153	1	MYG_LAGLA
40	6	5.7	153	1	MYG_LAGMA
41	6	5.7	153	1	MYG_LEPMU
42	6	5.7	153	1	MYG_LOXAF
43	6	5.7	153	1	MYG_LUTLU
44	6	5.7	153	1	MYG_LYCP1
45	6	5.7	153	1	MYG_MACRA
46	6	5.7	153	1	MYG_MEGNO
47	6	5.7	153	1	MYG_MEIME
48	6	5.7	153	1	MYG_MESCA
49	6	5.7	153	1	MYG_NYCCO
50	6	5.7	153	1	MYG_OCHRR
51	6	5.7	153	1	MYG_ONDZ1
52	6	5.7	153	1	MYG_ORYAP
53	6	5.7	153	1	MYG_PANTR
54	6	5.7	153	1	MYG_PAPAN
55	6	5.7	153	1	MYG_PERPO
56	6	5.7	153	1	MYG_PHOPH
57	6	5.7	153	1	MYG_PHOS1
58	6	5.7	153	1	MYG_PIG
59	6	5.7	153	1	MYG_PONPY
60	6	5.7	153	1	MYG_PROGU
61	6	5.7	153	1	MYG_ROUNE
62	6	5.7	153	1	MYG_SAIAC
63	6	5.7	153	1	MYG_SHEEP
64	6	5.7	153	1	MYG_SPA8H
65	6	5.7	153	1	MYG_TUPGL
66	6	5.7	153	1	MYG_VULCH
67	6	5.7	153	1	MYG_ZALCA
68	6	5.7	153	1	MYG_ZIPCA
69	6	5.7	178	1	HPRT_BUCAT
70	6	5.7	179	1	YEEL_SALTY
71	6	5.7	181	1	GSPH_ERMCH
72	6	5.7	197	1	RASB_DICD1
73	6	5.7	201	1	RBCR_RHIOA
74	6	5.7	220	1	RP29_HUMAN
75	6	5.7	229	1	RS3_ARCFU
76	6	5.7	235	1	Y680_RHIOA
77	6	5.7	242	1	ATPE_RICPR
78	6	5.7	252	1	YDIY_ECOLI
79	6	5.7	260	1	MCT1_BAR
80	6	5.7	263	1	UCPA_ECO57
81	6	5.7	263	1	UCPA_ECOLI
82	6	5.7	263	1	UCPA_SALTY
83	6	5.7	269	1	PRTM_BACSP
84	6	5.7	269	1	SUBS_BACLE
85	6	5.7	269	1	SUBS_BACLE
86	6	5.7	301	1	PTB_CLOAB
87	6	5.7	322	1	Y268_BACDH
88	6	5.7	325	1	TDG_SCHPO
89	6	5.7	330	1	MTHT_MERTP
90	6	5.7	331	1	YOO1_CAREL
91	6	5.7	334	1	CHMU_ARATH
92	6	5.7	340	1	MPCP_CAREL
93	6	5.7	342	1	PS8_MOUSE
94	6	5.7	345	1	CATD_PIG
95	6	5.7	355	1	S18A_MOUSE
96	6	5.7	357	1	MTDC_DROME
97	6	5.7	365	1	CATD_SHEEP
98	6	5.7	372	1	ALR_SYNY3
99	6	5.7	373	1	YOO3_CAREL
100	6	5.7	374	1	G3PA_CHIRE
101	6	5.7	376	1	S18E_HUMAN
102	6	5.7	377	1	YD80_AGR75
103	6	5.7	380	1	ELYA_BACAO
104	6	5.7	380	1	ELYA_BACCS
105	6	5.7	384	1	AMPC_PROST
106	6	5.7	386	1	SYT5_HUMAN

P02162	halichoerus
P02168	equus caball
P02144	homo sapien
P02146	hylobates a
P02181	inla geoffr
P02154	lagochrix l
P02160	leptotomus
P02165	leptilemur m
P02187	loxodonta a
P11343	lutra lutra
P02159	lycaon pict
P02150	macaca fasc
P02178	megaptora n
P02157	meles meles
P02183	mesopiodon
P02167	mytilus edulis
P02171	ochotona pr
P32428	ondatra zib
P02164	oryzctopus
P02145	pan troglod
P02149	papio anubi
P02166	perodicticu
P02176	phococenoide
P30562	pioca sibir
P02189	sus scrofa
P02148	pongo pygma
P02163	proechimys
P02165	rousettus a
P02155	salimiri sci
P02190	ovis aries
P04248	spalax leuc
P02165	tupaiia glis
P02160	volpes cham
P02161	zalophus ca
P02182	ziphius cav
P57291	buchnera ap
P30196	salmonella
P24687	erwinia chr
P32255	dictyostel
Q98bm4	rhizobium l
Q95707	homo sapien
Q08360	archaeoglob
Q87440	escherichia
P37441	salmonella
Q99405	bacillus sp
P39599	bacillus le
P29600	bacillus le
P58255	clostridium
Q8K706	bacillus ha
Q59825	escheriasch
P29567	methanobact
P42463	caenorhabdi
P242758	arabidopsis
P40614	caenorhabdi
Q96312	chlamydomon
Q15466	homo sapien
Q8uf17	agrobacteri
P27693	bacillus al
P41362	bacillus ci
Q69773	providencia
Q00445	homo sapien









545	5	4.7	228	1	YMEB_METEX	Q99116 methylobact	618	5	4.7	251	1	PCN2_AERPE	Q9977 aeropyrum p
546	5	4.7	229	1	GLC8_YEAST	P41818 saccharomyc	619	5	4.7	252	1	PMG2_HUMAN	P15259 homo sapien
547	5	4.7	229	1	MTN_HAEN	P45113 haemophilus	620	5	4.7	253	1	KDUD_ECOLI	P17769 escherichia
548	5	4.7	230	1	PCRB_STAM	O991n1 staphylococ	621	5	4.7	253	1	KDUD_ERWCH	O05528 erwina chr
549	5	4.7	230	1	PCRB_STAM	O995y1 staphylococ	622	5	4.7	253	1	NOCT_RAT	O9et55 rattus norv
550	5	4.7	230	1	RK1_ODOST	P49544 odonella s	623	5	4.7	253	1	POMA_VIBAL	O06873 vibrio algi
551	5	4.7	231	1	ADPE_YEAST	Q01976 saccharomyc	624	5	4.7	253	1	RS4E_AERPE	Q9915 aeropyrum p
552	5	4.7	231	1	ALSE_ECOLI	P37719 escherichia	625	5	4.7	253	1	TPIS_SCHMA	P48501 schistosoma
553	5	4.7	231	1	FLGD_ECOLI	P75936 escherichia	626	5	4.7	253	1	YU6E_DEIRA	Q97c3 diinococcus
554	5	4.7	231	1	PSBO_ONOVI	O22591 onobrychis	627	5	4.7	255	1	2BHD_STREX	P19992 streptomyce
555	5	4.7	231	1	VA11_MYCPN	P75102 mycoplasma	628	5	4.7	255	1	FABG_MYCAV	O07399 mycobacteri
556	5	4.7	232	1	FLGD_SALT	P16321 salmoneilla	629	5	4.7	255	1	FLAS_PYRHO	O58286 pyrococcus
557	5	4.7	233	1	VARP_LAMB	P03689 bacteriopho	630	5	4.7	255	1	RPSF_BACSU	P07860 bacillus su
558	5	4.7	234	1	Y340_METUA	O57786 methanococc	631	5	4.7	255	1	TNR9_HUMAN	Q07011 homo sapien
559	5	4.7	235	1	YDJ2_ECOLI	P76221 escherichia	632	5	4.7	255	1	VG54_BPMD2	O64245 mycobacteri
560	5	4.7	235	1	YIP2_YEAST	P40455 saccharomyc	633	5	4.7	256	1	PDX3_HUMAN	P34067 homo sapien
561	5	4.7	236	1	COAT_PXYHB	O07626 poteco viru	634	5	4.7	256	1	PDX3_HUMAN	P34067 homo sapien
562	5	4.7	237	1	GLDB_STRPY	Q991d7 streptococc	635	5	4.7	257	1	TRIE_DROER	P54677 drosophila
563	5	4.7	237	1	MINC_BUCAI	P57412 buchnera ap	636	5	4.7	257	1	FMBX_BACNO	P17834 bacteroides
564	5	4.7	237	1	RECA_NEIDE	O86403 neisseria ap	637	5	4.7	257	1	HS70_LUPPO	P27905 bacteroides
565	5	4.7	237	1	RECA_NEISI	O86411 neisseria s	638	5	4.7	257	1	NUDC_ECO57	P16121 lupinus pol
566	5	4.7	237	1	RECA_NEISU	O86414 neisseria s	639	5	4.7	257	1	NUDC_ECOLI	O8x647 escherichia
567	5	4.7	237	1	RK1_CYARA	P48125 cyanophora	640	5	4.7	257	1	NUDC_SALTI	P32664 escherichia
568	5	4.7	237	1	Y751_TREPA	O83732 treponema p	641	5	4.7	257	1	NUDC_SALTY	O82328 salmoneilla
569	5	4.7	237	1	YG6C_ECOLI	P11664 escherichia	642	5	4.7	257	1	PDX3_BOVIN	Q91515 salmoneilla
570	5	4.7	238	1	Y538_CHLTR	O84543 chlamydia t	643	5	4.7	257	1	PDX3_MOUSE	P315705 bos taurus
571	5	4.7	238	1	Y825_CHLMU	O991k4 chlamydia m	644	5	4.7	257	1	RL7A_ARATH	P49692 arabidopsis
572	5	4.7	239	1	HIS4_MERTH	O26993 methanobact	645	5	4.7	258	1	LPXA_NRIAB	P95379 neisseria m
573	5	4.7	240	1	HIS4_LISIN	Q92e87 listeria in	646	5	4.7	258	1	RNH2_LACTA	Q96917 lactococcus
574	5	4.7	241	1	AGL9_PETHY	Q03489 petunia hyb	647	5	4.7	259	1	LGT_STRMU	P72462 streptococc
575	5	4.7	241	1	VMOM_BPMU	P06018 bacteriopho	648	5	4.7	259	1	Y714_STREH	P20182 streptomyce
576	5	4.7	241	1	Y708_VIBCH	O9ku21 vibrio chol	649	5	4.7	260	1	1433_OENHO	P29307 oenothera h
577	5	4.7	242	1	RRPO_SCVLA	P23172 saccharomyc	650	5	4.7	260	1	AG22_SHEEP	Q28999 ovis aries
578	5	4.7	242	1	RS4E_MERTH	O26123 methanobact	651	5	4.7	260	1	CY12_SOLTU	P29610 solanum tub
579	5	4.7	242	1	TRPD_BACCA	P30525 bacillus ca	652	5	4.7	260	1	GLYA_CORSI	P50434 corynebacte
580	5	4.7	243	1	RADC_SYNT3	P52601 synechocyst	653	5	4.7	260	1	NUDC_YERPE	O8245 yersteinia pe
581	5	4.7	243	1	T2N4_NEILA	P50183 neisseria 1	654	5	4.7	260	1	YK02_MYCTU	Q10855 mycobacteri
582	5	4.7	243	1	T0NB_KLEBN	P45014 klebsiella	655	5	4.7	260	1	YRBE_ECOLI	P45332 escherichia
583	5	4.7	243	1	Y912_METUA	O58322 methanococc	656	5	4.7	261	1	PCNA_CRIGR	P57761 cricetus
584	5	4.7	244	1	COAT_MSVK	P03569 maize strea	657	5	4.7	261	1	PCNA_HUMAN	P12004 homo sapien
585	5	4.7	244	1	COAT_MSVN	P06448 maize strea	658	5	4.7	261	1	PCNA_MOUSE	P17918 mus musculu
586	5	4.7	244	1	COAT_MSVS	P14986 maize strea	659	5	4.7	261	1	PCNA_RAT	P04961 rattus norv
587	5	4.7	244	1	CSMI_CHLTE	O68988 chlorobium	660	5	4.7	262	1	VATD_ARATH	Q9xgm1 arabidopsis
588	5	4.7	244	1	MCT2_MOUSE	P51119 mus musculu	661	5	4.7	262	1	YXBG_BACSU	P46311 bacillus su
589	5	4.7	244	1	RS4E_METUA	P14023 methanococc	662	5	4.7	263	1	ACON_SRRMU	O59938 streptococc
590	5	4.7	245	1	C1OC_HUMAN	P02747 homo sapien	663	5	4.7	263	1	YG11_PSEPU	P31866 pseudomonas
591	5	4.7	246	1	HEM4_SALTY	Q916q3 salmoneilla	664	5	4.7	264	1	CB23_APTGR	P92919 apium grave
592	5	4.7	246	1	MCT1_MERUN	P50340 meriones un	665	5	4.7	264	1	YES4_SCHPO	O14171 schizosach
593	5	4.7	246	1	PHBB_CHRYI	P56728 thermotoga	666	5	4.7	264	1	ELH2_ACICA	O26659 methanobact
594	5	4.7	246	1	Y928_THEMA	O50673 mycobacteri	667	5	4.7	266	1	Y579_MERTH	Q9ck9 pasteurella
595	5	4.7	246	1	Y928_THEMA	O50673 mycobacteri	668	5	4.7	267	1	CYS2_PASMU	O44023 alcaligenes
596	5	4.7	247	1	COMB_CLOPE	Q8xhc3 clostridium	669	5	4.7	267	1	LEU2_ALCEU	Q9zm66 helicobacte
597	5	4.7	247	1	DB83_HUMAN	P57088 homo sapien	670	5	4.7	270	1	PANB_HELPJ	P03667 bacteriopho
598	5	4.7	247	1	DB83_RAT	Q9z142 rattus norv	671	5	4.7	271	1	VG18_BPP22	O88666 mus musculu
599	5	4.7	247	1	FABG_CHLTR	P38004 chlamydia t	672	5	4.7	272	1	CLPP_MOUSE	P52528 mycoplasma
600	5	4.7	247	1	FABG_MYCTU	O48930 mycobacteri	673	5	4.7	272	1	ESL3_MYCPN	P58209 escherichia
601	5	4.7	247	1	KDSB_ECO57	O8x66 escherichia	674	5	4.7	273	1	DABP_ECO57	P52076 notophthal
602	5	4.7	247	1	KDSB_ECOLI	P04993 escherichia	675	5	4.7	273	1	DABP_ECOLI	P53703 helicobacte
603	5	4.7	247	1	KDSB_SALTI	Q82800 salmoneilla	676	5	4.7	273	1	DLX3_NOTVI	O29m77 helicobacte
604	5	4.7	247	1	KDSB_SALTY	Q8xgc2 salmoneilla	677	5	4.7	273	1	FLJL_CAUCR	O24990 homo sapien
605	5	4.7	248	1	MTAA_SYNP2	P34882 synechococc	678	5	4.7	274	1	DLX3_PLEWA	P52399 clostridium
606	5	4.7	248	1	NDDB_MOUSE	O35417 mus musculu	679	5	4.7	274	1	FRTI_MOUSE	P09303 pleurodeles
607	5	4.7	248	1	NDDB_RAT	P06300 rattus norv	680	5	4.7	274	1	RECA_NEISU	P70393 mus musculu
608	5	4.7	248	1	PHAB_ACTIS	P50203 acinetobact	681	5	4.7	274	1	RECA_NEIMU	O59577 neisseria f
609	5	4.7	248	1	VD10_VAUV	P33071 variola vir	682	5	4.7	275	1	PABI_HELPJ	O59619 neisseria m
610	5	4.7	249	1	BTUD_SALTY	Q8xgf8 salmoneilla	683	5	4.7	275	1	PABI_HELPJ	O29m77 helicobacte
611	5	4.7	249	1	MCT1_CANPA	P21842 canis fami1	684	5	4.7	277	1	CLPP_HUMAN	O16740 homo sapien
612	5	4.7	250	1	AGL9_ARADE	O38694 aranda debo	685	5	4.7	277	1	SP0K_CLOIN	P52399 clostridium
613	5	4.7	250	1	DABP_STRCO	P10682 streptomyce	686	5	4.7	278	1	IE66_VZVD	P09355 varicella-z
614	5	4.7	250	1	HXP4_MOUSE	P15394 agrobacteri	687	5	4.7	278	1	PROI_BACSU	P54552 bacillus su
615	5	4.7	250	1	REPA_AGRTU	Q9y910 aeropyrum p	688	5	4.7	278	1	T2P2_NEIGO	P24617 neisseria g
616	5	4.7	250	1	TDXH_AERPE	Q10405 mycobacteri	689	5	4.7	278	1	UPK2_AGRTS	P58741 agrobacteri
617	5	4.7	250	1	YMI9_MYCTU		690	5	4.7	278	1	UPK2_SUNSO	Q97494 sulfolobus





Accession	Length	Score	DB 1	Length	DB 2	Indels	Gaps
P31216 escherichia	28	264	CERCARIAL PROTEASE.	28	264	0	0
P45767 escherichia	68	68	CHARGE RELAY SYSTEM (BY SIMILARITY).	68	68	0	0
P35249 homo sapien	126	126	CHARGE RELAY SYSTEM (BY SIMILARITY).	126	126	0	0
P50052 homo sapien	218	218	CHARGE RELAY SYSTEM (BY SIMILARITY).	218	218	0	0
Q92026 meriones un	264	264	CHARGE RELAY SYSTEM (BY SIMILARITY).	264	264	0	0
P35374 mus musculu	28545	28545	CHARGE RELAY SYSTEM (BY SIMILARITY).	28545	28545	0	0
P32335 schizosacch	100.0%	100.0%	CHARGE RELAY SYSTEM (BY SIMILARITY).	100.0%	100.0%	0	0
Q95884 schizosacch	100.0%	100.0%	CHARGE RELAY SYSTEM (BY SIMILARITY).	100.0%	100.0%	0	0
Q97493 thermoplas	100.0%	100.0%	CHARGE RELAY SYSTEM (BY SIMILARITY).	100.0%	100.0%	0	0
Q59110 archaeoglob	100.0%	100.0%	CHARGE RELAY SYSTEM (BY SIMILARITY).	100.0%	100.0%	0	0
Q9pcfs xylella fas	100.0%	100.0%	CHARGE RELAY SYSTEM (BY SIMILARITY).	100.0%	100.0%	0	0
Q01017 herpesvirus	100.0%	100.0%	CHARGE RELAY SYSTEM (BY SIMILARITY).	100.0%	100.0%	0	0

## ALIGNMENTS

RESULT 1

Accession	Length	Score	DB 1	Length	DB 2	Indels	Gaps
P31216 escherichia	28	264	CERCARIAL PROTEASE.	28	264	0	0
P45767 escherichia	68	68	CHARGE RELAY SYSTEM (BY SIMILARITY).	68	68	0	0
P35249 homo sapien	126	126	CHARGE RELAY SYSTEM (BY SIMILARITY).	126	126	0	0
P50052 homo sapien	218	218	CHARGE RELAY SYSTEM (BY SIMILARITY).	218	218	0	0
Q92026 meriones un	264	264	CHARGE RELAY SYSTEM (BY SIMILARITY).	264	264	0	0
P35374 mus musculu	28545	28545	CHARGE RELAY SYSTEM (BY SIMILARITY).	28545	28545	0	0
P32335 schizosacch	100.0%	100.0%	CHARGE RELAY SYSTEM (BY SIMILARITY).	100.0%	100.0%	0	0
Q95884 schizosacch	100.0%	100.0%	CHARGE RELAY SYSTEM (BY SIMILARITY).	100.0%	100.0%	0	0
Q97493 thermoplas	100.0%	100.0%	CHARGE RELAY SYSTEM (BY SIMILARITY).	100.0%	100.0%	0	0
Q59110 archaeoglob	100.0%	100.0%	CHARGE RELAY SYSTEM (BY SIMILARITY).	100.0%	100.0%	0	0
Q9pcfs xylella fas	100.0%	100.0%	CHARGE RELAY SYSTEM (BY SIMILARITY).	100.0%	100.0%	0	0
Q01017 herpesvirus	100.0%	100.0%	CHARGE RELAY SYSTEM (BY SIMILARITY).	100.0%	100.0%	0	0

RESULT 2

Accession	Length	Score	DB 1	Length	DB 2	Indels	Gaps
P31216 escherichia	28	264	CERCARIAL PROTEASE.	28	264	0	0
P45767 escherichia	68	68	CHARGE RELAY SYSTEM (BY SIMILARITY).	68	68	0	0
P35249 homo sapien	126	126	CHARGE RELAY SYSTEM (BY SIMILARITY).	126	126	0	0
P50052 homo sapien	218	218	CHARGE RELAY SYSTEM (BY SIMILARITY).	218	218	0	0
Q92026 meriones un	264	264	CHARGE RELAY SYSTEM (BY SIMILARITY).	264	264	0	0
P35374 mus musculu	28545	28545	CHARGE RELAY SYSTEM (BY SIMILARITY).	28545	28545	0	0
P32335 schizosacch	100.0%	100.0%	CHARGE RELAY SYSTEM (BY SIMILARITY).	100.0%	100.0%	0	0
Q95884 schizosacch	100.0%	100.0%	CHARGE RELAY SYSTEM (BY SIMILARITY).	100.0%	100.0%	0	0
Q97493 thermoplas	100.0%	100.0%	CHARGE RELAY SYSTEM (BY SIMILARITY).	100.0%	100.0%	0	0
Q59110 archaeoglob	100.0%	100.0%	CHARGE RELAY SYSTEM (BY SIMILARITY).	100.0%	100.0%	0	0
Q9pcfs xylella fas	100.0%	100.0%	CHARGE RELAY SYSTEM (BY SIMILARITY).	100.0%	100.0%	0	0
Q01017 herpesvirus	100.0%	100.0%	CHARGE RELAY SYSTEM (BY SIMILARITY).	100.0%	100.0%	0	0

RESULT 3

Accession	Length	Score	DB 1	Length	DB 2	Indels	Gaps
P31216 escherichia	28	264	CERCARIAL PROTEASE.	28	264	0	0
P45767 escherichia	68	68	CHARGE RELAY SYSTEM (BY SIMILARITY).	68	68	0	0
P35249 homo sapien	126	126	CHARGE RELAY SYSTEM (BY SIMILARITY).	126	126	0	0
P50052 homo sapien	218	218	CHARGE RELAY SYSTEM (BY SIMILARITY).	218	218	0	0
Q92026 meriones un	264	264	CHARGE RELAY SYSTEM (BY SIMILARITY).	264	264	0	0
P35374 mus musculu	28545	28545	CHARGE RELAY SYSTEM (BY SIMILARITY).	28545	28545	0	0
P32335 schizosacch	100.0%	100.0%	CHARGE RELAY SYSTEM (BY SIMILARITY).	100.0%	100.0%	0	0
Q95884 schizosacch	100.0%	100.0%	CHARGE RELAY SYSTEM (BY SIMILARITY).	100.0%	100.0%	0	0
Q97493 thermoplas	100.0%	100.0%	CHARGE RELAY SYSTEM (BY SIMILARITY).	100.0%	100.0%	0	0
Q59110 archaeoglob	100.0%	100.0%	CHARGE RELAY SYSTEM (BY SIMILARITY).	100.0%	100.0%	0	0
Q9pcfs xylella fas	100.0%	100.0%	CHARGE RELAY SYSTEM (BY SIMILARITY).	100.0%	100.0%	0	0
Q01017 herpesvirus	100.0%	100.0%	CHARGE RELAY SYSTEM (BY SIMILARITY).	100.0%	100.0%	0	0

Db 168 SGIRVISTL 175

## RESULT 3

TRYP\_STRGA STANDARD; PRT; 268 AA.  
ID TRYP\_STRGA  
AC Q54179;  
DT 01-NOV-1997 (Rel. 35, Created)  
DT 01-NOV-1997 (Rel. 35, Last sequence update)  
DT 15-JUN-2002 (Rel. 41, Last annotation update)  
DE Trypsin-like protease precursor (EC 3.4.21.-).  
OS Streptomyces glaucescens.  
OC Bacteria; Actinobacteria; Actinobacteridae;  
OC Actinomycetales; Streptomycinae; Streptomycetaceae; Streptomyces.  
NX NCBI\_TaxId=1907;  
[1]  
RF SEQUENCE FROM N.A.  
RC STRAIN=ATCC 13032 / ETH 22794 / GLA.0;  
RA Hintemann G.;  
RT "Cloning, partial characterization and nucleotide sequence of a trypsin-like protease gene from Streptomyces glaucescens."  
RL Submitted (Aug-1994) to the EMBL/Genbank/DBJ databases.  
CC -1- CATALYTIC ACTIVITY: Preferential cleavage: Arg-, Lys-.  
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL, U13770; AAA21336.1; -.  
DR HSSP; P00775; 1SGT.  
DR MEROPS; S01.101; -.  
DR InterPro; IPR001314; Chymotrypsin.  
DR InterPro; IPR001254; Ser protease\_Try.  
DR Pfam; PF00089; trypsin\_1.  
DR PRINTS; PR00722; CHYMOTRYPSIN.  
DR SMART; SM00020; Tryp\_Spc; 1.  
DR PROSITE; PS50240; TRYPSIN\_DOM; 1.  
DR PROSITE; PS00134; TRYPSIN\_HIS; 1.  
DR PROSITE; PS00135; TRYPSIN\_SER; 1.  
KM Hydroxylase; Serine protease; Zymogen; Signal.  
FT SIGNAL 1 41  
FT PROPEP 42 45  
FT CHAIN 46 268  
FT ACT\_SITE 82 82 TRYPSIN-LIKE PROTEASE.  
FT ACT\_SITE 127 127 CHARGE RELAY SYSTEM (BY SIMILARITY).  
FT ACT\_SITE 217 217 CHARGE RELAY SYSTEM (BY SIMILARITY).  
FT DISULFID 67 83  
FT DISULFID 187 202  
FT DISULFID 213 242  
FT SITE 211 211 REQUIRED FOR SPECIFICITY (BY SIMILARITY).  
SQ SEQUENCE 268 AA; 27550 MW; 01D05E0A4D5D177A CRC64;

Query Match 6.6%; Score 7; DB 1; Length 268;  
Best Local Similarity 100.0%; Pred. No. 8;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 21 VKVAPGY 27  
Db 115 VKVAPGY 121

RESULT 4  
ID LDBH\_RHIOH STANDARD; PRT; 302 AA.  
AC Q9P4B5;  
DT 15-JUN-2002 (Rel. 41, Created)  
DT 15-JUN-2002 (Rel. 41, Last sequence update)

DT 15-JUN-2002 (Rel. 41, Last annotation update)  
DE L-lactate dehydrogenase B (EC 1.1.1.27) (L-LDH B).  
GN LDBH.  
OS Rhizopus oryzae (Rhizopus delemar).  
OC Eukaryota; Fungi; Zygomycota; Zygomycetes; Mucorales; Mucoraceae;  
OC Rhizopus.  
NX NCBI\_TaxId=64495;  
[1]  
RF SEQUENCE FROM N.A.  
RC STRAIN=NRRL 395;  
RX MEDLINE=20292801; PubMed=10831409;  
RA Skory C.D.;  
RT "Isolation and expression of lactate dehydrogenase genes from Rhizopus  
RT oryzae."  
RL Appl. Environ. Microbiol. 66:2343-2348(2000).  
CC -1- FUNCTION: Converts L-lactate to pyruvate.  
CC -1- CATALYTIC ACTIVITY: (S)-lactate + NAD(+) = pyruvate + NADH.  
CC -1- PATHWAY: Anaerobic glycolysis; final step.  
CC -1- SUBUNIT: Homotetramer (By similarity).  
CC -1- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).  
CC -1- SIMILARITY: BELONGS TO THE LDH FAMILY.  
CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL; AF226155; AAF61914.1; -.  
DR HSSP; P00344; 1LDB.  
DR InterPro; IPR001557; L\_LDH.  
DR InterPro; IPR000205; NAD\_binding.  
DR InterPro; IPR001236; 1dh.  
DR Pfam; PF00056; 1dh; 1.  
DR Pfam; PF02866; 1dh\_C; 1.  
DR PRINTS; PR00086; 1LDHHDGNASE.  
DR PROSITE; PS00064; L\_LDH; 1.  
KM Oxidoreductase; NAD; Glycolysis  
FT ACT\_SITE 175 175 ACCEPTS A PROTON DURING CATALYSIS  
FT SITE 175 175 (BY SIMILARITY).  
SQ SEQUENCE 302 AA; 32387 MW; 348C79DF8ABDE695 CRC64;

Query Match 6.6%; Score 7; DB 1; Length 302;  
Best Local Similarity 100.0%; Pred. No. 8.8;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 40 IAQTLSG 46  
Db 128 IAQTLSG 134

## RESULT 5

ID S18A\_HUMAN STANDARD; PRT; 356 AA.  
AC Q92185; Q93064;  
DT 01-NOV-1997 (Rel. 35, Created)  
DT 01-NOV-1997 (Rel. 35, Last sequence update)  
DT 15-JUN-2002 (Rel. 41, Last annotation update)  
DE Alpha-N-acetyl-neuraminidase alpha-2,8-sialyltransferase (EC 2.4.99.8)  
DE (Ganglioside GD3 synthase) (Ganglioside Gm3 synthase) (Alpha-2,8-sialyltransferase 8A) (ST8SIA1).  
GN ST8SIA OR ST8A.  
OS Homo sapiens (human).  
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
OX NCBI\_TaxId=9606;  
[1]  
RF SEQUENCE FROM N.A.  
RX MEDLINE=95024133; PubMed=7937974;  
RA Haraguchi M., Yamashiro S., Yamamoto A., Furukawa K., Takamiya K., Lloyd K.O., Shiku H., Furukawa K.;

```

RT "Isolation of GD3 synthase gene by expression cloning of GM3
RT alpha-2,8-sialyltransferase cDNA using anti-GD2 monoclonal
RT antibody."
RL Proc. Natl. Acad. Sci. U.S.A. 91:10455-10459(1994).
[2]
RN SEQUENCE FROM N.A.
RP MEDLINE=9433665; PubMed=8058740;
RA Nara K., Watanabe Y., Maruyama K., Kasahara K., Nagai Y., Sanai Y.,
RT "Expression cloning of a CMP-NeuAc:NeuAc alpha 2-3Gal beta 1-4Glc
RT beta 1-1'Cer alpha 2,8-sialyltransferase (GD3 synthase) from human
RT melanoma cells."
RL Proc. Natl. Acad. Sci. U.S.A. 91:7952-7956(1994).
[3]
RN SEQUENCE FROM N.A.
RP MEDLINE=96216472; PubMed=8631981;
RA Nakayama J., Fukuda M.N., Hirabayashi Y., Kanamori A., Sasaki K.,
RT "Expression cloning of a human GT3 synthase. GD3 and GT3 are
RT synthesized by a single enzyme."
RL J. Biol. Chem. 271:3684-3691(1996).
[4]
RN SEQUENCE FROM N.A.
RP MEDLINE=94253194; PubMed=8195250;
RA Sasaki K., Kurata K., Kojima N., Kurosawa N., Ohta S., Hanai N.,
RT Tsuji S., Nishi T.,
RT "Expression cloning of a GM3-specific alpha-2,8-sialyltransferase
RT (GD3 synthase)."
RL J. Biol. Chem. 269:15950-15956(1994).
CC -I- FUNCTION: Involved in the production of GD3 and GT3 from GM3.
CC -I- CATALYTIC ACTIVITY: CMP-N-acetylneuraminate + alpha-N-
CC acetylneuraminy-1-2,3-beta-D-galactosyl-R = CMP + alpha-N-
CC acetylneuraminy-1-2,8-alpha-N-acetylneuraminy-1-2,3-beta-D-
CC galactosyl-R.
CC -I- PATHWAY: Glycosylation.
CC -I- SUBCELLULAR LOCATION: Type II membrane protein. Golgi (potential).
CC -I- TISSUE SPECIFICITY: STRONGLY EXPRESSED IN MELANOMA CELL LINES,
CC ADULT AND FETAL BRAIN AND TO A LESSER EXTENT IN ADULT AND FETAL
CC LUNG.
CC -I- SIMILARITY: BELONGS TO THE GLYCOSYLTRANSFERASE FAMILY 29.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
CC EMBL; L32867; AAA62366.1; -.
CC EMBL; D26360; BAA05391.1; -.
CC EMBL; L43494; AAC37586.1; ALT_INIT.
CC EMBL; X77922; CA54891.1; ALT_INIT.
CC Genew; HGNC:10869; SIRT9A.
CC MIM; 601123; -.
CC DR InterPro; IPR001675; GT_29.
CC DR Pfam; PF00777; Glyco_transf_29; 1.
CC KW Transferase; Glycosyltransferase; Glycoprotein; Transmembrane;
CC Signal-anchor; Golgi stack.
CC FT DOMAIN 1 29 CYTOPLASMIC (POTENTIAL).
CC TRANSMEM 30 48 SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN)
CC (POTENTIAL).
CC FT DOMAIN 49 356 LUMENAL, CATALYTIC (POTENTIAL).
CC FT CARBOHYD 71 71 N-LINKED (GLCNAC. . .) (POTENTIAL).
CC FT CARBOHYD 119 119 N-LINKED (GLCNAC. . .) (POTENTIAL).
CC FT CARBOHYD 214 214 N-LINKED (GLCNAC. . .) (POTENTIAL).
CC FT CARBOHYD 245 245 N-LINKED (GLCNAC. . .) (POTENTIAL).
CC SEQUENCE 356 AA; 40519 MW; 452FE04856964395 CRC64;
Query Match 6.6%; Score 7; DB 1; Length 356;
Best Local Similarity 100.0%; Pred. No. 10;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Oy 100 NGILKK 106

```

```

Db 143 NGILKK 149
|||||
RESULT 6
MTDI_NEIMC STANDARD; PRT; 420 AA.
ID MTDI_NEIMC
AC O9RLM4;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Probable modification methylase NmeDIP (EC 2.1.1.73) (Cytosine-
DE specific methyltransferase NmeDIP) (M.NmeDIP).
GN NmeDIP.
OS Neisseria meningitidis (serogroup C).
OC Bacteria; Proteobacteria; beta subdivision; Neisseriaceae; Neisseria.
OX NCBI_TaxID=135720;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=2120 / Serogroup C / Serotype NT;
RX MEDLINE=20138154; PubMed=10671450;
RA Claus H., Friedrich A., Frosch M., Vogel U.;
RT "Differential distribution of novel restriction-modification systems
RT in clonal lineages of Neisseria meningitidis."
RL J. Bacteriol. 182:1296-1303(2000).
CC -I- FUNCTION: THIS METHYLASE RECOGNIZES A DOUBLE-STRANDED SEQUENCE,
CC CAUSES SPECIFIC METHYLATION ON A C ON BOTH STRANDS, AND PROTECTS
CC THE DNA FROM CLEAVAGE BY THE NMEIDIP ENDONUCLEASE (BY SIMILARITY).
CC -I- CATALYTIC ACTIVITY: S-adenosyl-L-methionine + DNA cytosine = S-
CC adenosyl-L-homocysteine + DNA 5-methylcytosine.
CC -I- SIMILARITY: BELONGS TO THE C5-METHYLTRANSFERASE FAMILY.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
CC EMBL; AJ238948; CAB59897.1; -.
CC DR HSBP; P20589; IDCT.
CC DR REBASE; 4188; M.NmeDIP.
CC DR InterPro; IPR001525; C5_DNA_meth.
CC DR Pfam; PF00145; DNA_methylase; 1.
CC DR PRINTS; PR00105; C5METTRFAS.
CC DR TIGRFSMs; TIGR00675; dcm; 1.
CC DR PROSITE; PS00094; C5_MTASE_1; FALSE_NEG.
CC DR PROSITE; PS00095; C5_MTASE_2; FALSE_NEG.
CC KW Transferase; Methyltransferase; Restriction system.
CC FT ACT SITE 148 148 BY SIMILARITY.
CC SEQUENCE 420 AA; 47930 MW; 6E2472BA070354C9 CRC64;
Query Match 6.6%; Score 7; DB 1; Length 420;
Best Local Similarity 100.0%; Pred. No. 12;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Oy 44 LSGFDIA 50
|||||
Db 74 LSGFDIA 80
RESULT 7
PTR2_CANAL STANDARD; PRT; 623 AA.
ID PTR2_CANAL
AC P46030;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 01-NOV-1995 (Rel. 32, Last annotation update)
DE Peptide transporter PTR2.
GN PTR2.
OS Candida albicans (Yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;

```



CC Saccharomycetales; mitosporic Saccharomycetales; Candida.  
 OX NCBI\_TaxID=5476;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=CS 562 / ATCC 18804;  
 RX MEDLINE=95291458; PubMed=7773409;  
 RA Baerai M.A., Lubkowitz M.A., Perry J.R., Miller D., Krainer E.,  
 RA Nader F.R., Becker J.M.;  
 RT "Cloning of a Candida albicans peptide transport gene";  
 RL Microbiology 141:1147-1156(1995).  
 CC -1- FUNCTION: UPTAKE OF SMALL PEPTIDE.  
 CC -1- SUBCELLULAR LOCATION: Integral membrane protein.  
 CC -1- SIMILARITY: BELONGS TO THE PTR2 FAMILY OF TRANSPORTERS.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL, U09781, AAA80167.1, -.  
 DR InterPro, IPR00109, PTR2.  
 DR Pfam, PF00854, PTR2, 1.  
 DR PROSITE, PS01022; PTR2\_1; 1.  
 DR PROSITE, PS01023; PTR2\_2; 1.  
 KM Peptide transport; Transport; Transmembrane.  
 FT TRANSMEM 134 154 POTENTIAL.  
 FT TRANSMEM 163 183 POTENTIAL.  
 FT TRANSMEM 191 211 POTENTIAL.  
 FT TRANSMEM 250 270 POTENTIAL.  
 FT TRANSMEM 277 297 POTENTIAL.  
 FT TRANSMEM 385 405 POTENTIAL.  
 FT TRANSMEM 418 438 POTENTIAL.  
 FT TRANSMEM 448 468 POTENTIAL.  
 FT TRANSMEM 499 519 POTENTIAL.  
 FT TRANSMEM 529 549 POTENTIAL.  
 FT TRANSMEM 557 577 POTENTIAL.  
 SQ SEQUENCE 623 AA; 69941 MW; 1601FD3AE21B80EB CRC64;  
 Query Match 6.6%; Score 7; DB 1; Length 623;  
 Best Local Similarity 100.0%; Pred. No. 16;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 50 AIWMLAQ 56  
 DB 572 AIWMLAQ 578  
 RESULT 8  
 YK79\_MYCTU STANDARD; PRT; 656 AA.  
 ID YK79\_MYCTU  
 AC Q10687;  
 DT 01-OCT-1996 (Rel. 34, Created)  
 DT 01-OCT-1996 (Rel. 34, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 GN Hypothetical protein Rv2079.  
 GN Rv2079 OR MT2140 OR MTCY49.18.  
 OS Mycobacterium tuberculosis.  
 OC Bacteria; Actinobacteria; Actinobacteriales; Actinobacteridae;  
 OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.  
 OX NCBI\_TaxID=1773;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=H37Rv;  
 RX MEDLINE=98295987; PubMed=9634230;  
 RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,  
 RA Gordon S.V., Eiglmeier K., Gae S., Barry C.E. III, Tekala F.,  
 RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,  
 RA Davies R., Devlin K., Feltham D., Gentles S., Hamlin N., Holroyd S.,  
 RA Hornsby T., Jagals K., Krogh A., McLean J., Moule S., Murphy L.,  
 RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,

RA Rutter S., Seeger K., Skelton S., Squares S., Squares R.,  
 RA Sulston J.E., Taylor K., Whitehead S., Barrett B.G.;  
 RT "Deciphering the biology of Mycobacterium tuberculosis from the  
 RT complete genome sequence";  
 RL Nature 393:537-544(1998).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=CDC 1551 / Oshkosh;  
 RA Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,  
 RA Peterson J., DeBoy R., Dodson R., Gwinn M.L., Haft D., Hickey E.,  
 RA Kolony J.F., Nelson W.C., Umayan L.A., Esmolaev M.D., Salzberg S.L.,  
 RA Delcher A., Utterback T., Weidman J., Knouri H., Gill J., Mikula A.,  
 RA Bishai W.;  
 RT "Whole genome comparison of Mycobacterium tuberculosis clinical and  
 RT laboratory strains";  
 RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.  
 CC -1- SIMILARITY: SOME, TO M.TUBERCULOSIS Rv0963C.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL, Z73966, CA98192.1, -.  
 DR EMBL, AE007063, AAK46422.1, -.  
 DR TIGR, MT2140, -.  
 DR TubercuList, Rv2079, -.  
 KM Hypothetical protein; Complete proteome.  
 FT CONFLICT 47 47 Y -> C (IN REF. 2).  
 SQ SEQUENCE 656 AA; 69823 MW; 9DFB74A58809DB4 CRC64;  
 Query Match 6.6%; Score 7; DB 1; Length 656;  
 Best Local Similarity 100.0%; Pred. No. 17;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 75 IPPPGTG 81  
 DB 222 IPPPGTG 228  
 RESULT 9  
 MYT1\_HUMAN STANDARD; PRT; 1121 AA.  
 ID MYT1\_HUMAN  
 AC Q01538; Q9UPV2; O94922;  
 DT 01-JUL-1993 (Rel. 26, Created)  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 GN Myelin transcription factor 1 (MYT1) (MYT1) (Proteolipid protein  
 GN binding protein) (PLPBL).  
 GN MYT1 OR PLPBL OR MYT1 OR KIA0835 OR KIA1050.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Brain;  
 RX MEDLINE=99156230; PubMed=10048485;  
 RA Nagase T., Ishikawa K.-I., Suyama M., Kikuno R., Hirosewa M.,  
 RA Miyajima N., Tanaka A., Kotani H., Nomura N., Ohara O.,  
 RT "Prediction of the coding sequences of unidentified human genes. XII.  
 RT The complete sequences of 100 new cDNA clones from brain which code  
 RT for large proteins in vitro";  
 RL DNA Res. 5:355-364(1998).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=21638749; PubMed=11780052;  
 RA Deloukas P., Matthews L.H., Ashurst J., Burton J., Gilbert J.G.R.,  
 RA Jones M., Stavrides G., Almeida J.P., Babbage A.K., Bagguley C.L.,  
 RA Bailey J., Barlow K.F., Bates K.N., Beard L.M., Beare D.M.,

RA Beasley O.P., Bird C.P., Blakey S.E., Bridgeman A.M., Brown A.J.,  
 RA Buck D., Burrill W.D., Butler A.P., Carder C., Carter N.P.,  
 RA Chapman J.C., Clapp M., Clark G., Clark L.N., Clark S.Y., Clee C.M.,  
 RA Clegg S., Cobley V.E., Collier R.E., Connor R.E., Corby N.R.,  
 RA Ellington A.G., Franklin J.A., Fraser A., French L., Garner P.,  
 RA Giffiths D.V., Giffiths C., Giffiths M.N.D., Gwilliam R., Hall R.E.,  
 RA Hammond S., Harley J.L., Heath P.D., Ho S., Holden J.L., Howden P.J.,  
 RA Huckle E., Hunt A.R., Hunt S.E., Jekosch K., Johnson C.M., Johnson D.,  
 RA Kay M.P., Kimberley A.M., King A., Knights A., Laird G.K., Lawlor S.,  
 RA Levasseur M.H., Leversha M.A., Lloyd C., Lloyd D.M., Lovell J.D.,  
 RA Marsh V.A., Martin S.L., McCormachie L.J., McMay K., McKerrry A.A.,  
 RA Milne S.A., Mistry D., Moore M.J.F., Mullikin J.C., Nickerson T.,  
 RA Oliver K., Parker A., Patel R., Pearce T.A.V., Peck A.I.,  
 RA Phillimore B.J.C.T., Prathalingam S.R., Plumb R.W., Ramsay H.,  
 RA Rice C.M., Ross M.T., Scott C.E., Sehra H.K., Showkhen R., Sims S.,  
 RA Skuce C.D., Smith M.L., Soderlund C., Steward C.A., Sulston J.E.,  
 RA Swann R.M., Sycamore N., Taylor R., Tee L., Thomas D.W., Thorpe J.,  
 RA Tracey A., Tromans A.C., Vaudin M., Wall M., Wallis J.M.,  
 RA Whitehead S.L., Whitaker P., Willey D.L., Williams L., Williams S.A.,  
 RA Wilming L., Wray P.W., Hubbard T., Durbin R.M., Bentley D.R., Beck S.,  
 RA Rogers J.;  
 RT "The DNA sequence and comparative analysis of human chromosome 20.";  
 RL Nature 414:865-871(2001).  
 RN [3]  
 RP SEQUENCE OF 293-1121 FROM N.A.  
 RC TISSUE=Brain;  
 RX MEDLINE=99397452; PubMed=10470851;  
 RA Takano R., Nagase T., Ishikawa K.-I., Hirosewa M., Miyajima N.,  
 RA Tanaka A., Kotani H., Nomura N., Ohara O.;  
 RT "Prediction of the coding sequences of unidentified human genes. XIV.  
 RT The complete sequences of 100 new cDNA clones from brain which code  
 RT for large proteins in vitro.";  
 RL DNA Res. 6:197-205(1999).  
 RN [4]  
 RP SEQUENCE OF 424-1121 FROM N.A.  
 RC TISSUE=Brain;  
 RX MEDLINE=93078764; PubMed=1280325;  
 RA Kim J.G., Hudson L.D.;  
 RT "Novel member of the zinc finger superfamily: a C2-HC finger that  
 RT recognizes a glia-specific gene.";  
 RL Mol. Biol. 12:5632-5639(1992).  
 CC -1- FUNCTION: BINDS TO THE PROMOTER REGIONS OF PROTEOLIPID PROTEINS  
 CC OF THE CENTRAL NERVOUS SYSTEM.  
 CC -1- SUBCELLULAR LOCATION: Nuclear.  
 CC -1- TISSUE SPECIFICITY: MOSTLY IN DEVELOPING NERVOUS SYSTEM.  
 CC -1- DOMAIN: CONTAINS 7 ZINC FINGERS OF THE C2HC CLASS ARRANGED IN  
 CC CAN WIDELY SEPARATED CLUSTERS. THESE TWO DOMAINS OF DNA BINDING  
 CC FUNCTION INDEPENDENTLY AND RECOGNIZE THE SAME DNA SEQUENCE.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; AB030642; BAA74858.1; -;  
 DR EMBL; AL121581; CAC17005.1; -;  
 DR EMBL; AB028973; BAA83002.1; -;  
 DR EMBL; M96980; AAA59897.1; -;  
 DR PIR; A45033; A45033.  
 DR PIR; S27964; S27964.  
 DR TRANSFAC; T04937; -;  
 DR Genew; HGNC:7622; MYT1.  
 DR MIM; 600379; -;  
 DR InterPro; IPR002515; Znf\_C2HC.  
 DR Pfam; PF01530; zfc-C2HC; 7.  
 KW Transcription regulation; Zinc-finger; DNA-binding; Nuclear protein;  
 KW Repeat.  
 FT ZN\_FING 440 466 C2HC-TYPE.  
 FT ZN\_FING 484 510 C2HC-TYPE.

FT ZN\_FING 825 851 C2HC-TYPE.  
 FT ZN\_FING 874 895 C2HC-TYPE.  
 FT ZN\_FING 918 944 C2HC-TYPE.  
 FT ZN\_FING 971 997 C2HC-TYPE.  
 FT CONFLICT 567 574  
 FT CONFLICT 616 616  
 FT CONFLICT 837 837  
 FT CONFLICT 943 943  
 FT CONFLICT 957 958  
 FT CONFLICT 1075 1075  
 FT CONFLICT 1088 1089  
 FT CONFLICT 1102 1102  
 SQ SEQUENCE 1121 AA; 123239 MW; D4AF1F8C7D4EC01E CRC64;  
 Query Match 6.4%; Score 7; DB 1; Length 1121;  
 Best Local Similarity 100.0%; Pred. No. 26;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 Qy 19 SGVRYAP 25  
 |||||  
 Db 874 SGVRYAP 880  
 RESULT 10  
 CYAA\_NEUCR STANDARD; PRT; 2300 AA.  
 ID CYAA\_NEUCR  
 AC 001631;  
 DT 01-JUN-1993 (Rel. 26, Created)  
 DT 01-JUN-1993 (Rel. 26, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE Adenylate cyclase (EC 4.6.1.1) (ATP pyrophosphate-lyase) (adenyl1y1  
 DE cyclase).  
 GN CR-1 OR NAC.  
 OS Neurospora crassa.  
 CC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;  
 CC Sordariales; Sordariaceae; Neurospora.  
 OX NCBI\_TaxID=5141;  
 RX MEDLINE=9200795; PubMed=1680356;  
 RA Kore-Bda S., Murayama T., Uno I.;  
 RT "Isolation and characterization of the adenylate cyclase structural  
 RT gene of Neurospora crassa.";  
 RL Jpn. J. Genet. 66:317-334(1991).  
 CC -1- FUNCTION: PLAYS ESSENTIAL ROLES IN REGULATION OF CELLULAR  
 CC METABOLISM BY CATALYSING THE SYNTHESIS OF A SECOND MESSENGER,  
 CC CAMP.  
 CC -1- CATALYTIC ACTIVITY: ATP = 3',5'-cyclic AMP + diphosphate.  
 CC -1- SIMILARITY: BELONGS TO ADENYLYL CYCLASE CLASS-3 FAMILY.  
 CC -1- SIMILARITY: CONTAINS 23 LEUCINE-RICH REPEATS (LRR).  
 CC -1- SIMILARITY: CONTAINS 1 PP2C-LIKE DOMAIN.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; D00909; BAA00755.1; -;  
 DR InterPro; IPR001054; G\_cyclase.  
 DR InterPro; IPR001611; LRR.  
 DR InterPro; IPR003592; LRR\_out.  
 DR InterPro; IPR003591; LRR\_cyp.  
 DR InterPro; IPR001932; PP2C-like.  
 DR InterPro; IPR000159; RA domain.  
 DR Pfam; PF00211; guanylate\_cyc; 1.  
 DR Pfam; PF00481; PP2C; 1.  
 DR Pfam; PF00560; LRR; 14.  
 DR Pfam; PF00788; RA; 1.  
 DR PRINTS; PR00019; LEURICHRPT.

```

DR SMART; SM00044; CYC; 1.
DR SMART; SM00370; LRR; 4.
DR SMART; SM00369; LRR_TYP; 2.
DR SMART; SM00332; PP2C; 1.
DR SMART; SM00314; RA; 1.
DR PROSITE; PS0125; GUANYLATE_CYCLASES_2; 1.
KW Lyase; Repeat; leucine-rich repeat; CAMP synthesis; Magnesium.
FT REPEAT 867 890 LRR 1.
FT REPEAT 892 914 LRR 2.
FT REPEAT 915 938 LRR 3.
FT REPEAT 939 961 LRR 4.
FT REPEAT 962 986 LRR 5.
FT REPEAT 988 1008 LRR 6.
FT REPEAT 1009 1031 LRR 7.
FT REPEAT 1033 1055 LRR 8.
FT REPEAT 1056 1079 LRR 9.
FT REPEAT 1081 1097 LRR 10.
FT REPEAT 1098 1120 LRR 11.
FT REPEAT 1122 1142 LRR 12.
FT REPEAT 1143 1165 LRR 13.
FT REPEAT 1166 1188 LRR 14.
FT REPEAT 1189 1211 LRR 15.
FT REPEAT 1213 1234 LRR 16.
FT REPEAT 1349 1369 LRR 17.
FT REPEAT 1373 1396 LRR 18.
FT REPEAT 1398 1420 LRR 19.
FT REPEAT 1422 1445 LRR 20.
FT REPEAT 1447 1469 LRR 21.
FT REPEAT 1474 1497 LRR 22.
FT REPEAT 1513 1538 LRR 23.
FT DOMAIN 1564 1829 PP2C-LIKE.
FT DOMAIN 1830 2300 CATALYTIC.
FT DOMAIN 49 52 POLY-SER.
FT DOMAIN 204 221 POLY-GLN.
SQ SEQUENCE 2300 AA; 254752 MW; 52E79B90EBB17A7B CRC64;

Query Match
Best Local Similarity 100.0%; Score 7; DB 1; Length 2300;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 67 ISLPOPS 73
DB 346 ISLPOPS 352

RESULT 11
VXIS_BP434 STANDARD; PRT; 72 AA.
ID VXIS_BP434
AC P1683; P16408;
DT 01-OCT-1989 (Rel. 12, Created)
DT 01-AUG-1990 (Rel. 15, Last sequence update)
DT 01-OCT-1996 (Rel. 34, Last annotation update)
DE Excisionase.
GN XIS.
OS Bacteriophage 434, and
OS Bacteriophage HK022.
OC Viruses; dsDNA viruses, no RNA stage; Caudovirales; Siphoviridae;
OC Lambda-like viruses.
OX NCBI_TaxID=10712; 10742;
[1]
RP SEQUENCE FROM N.A.
RC SPECIES=Phage 434;
RX MEDLINE=91346141; PubMed=1715186;
RA Baker J., Limberger R., Schneider S.J., Campbell A.;
RT "Recombination and modular exchange in the genesis of new lambdaoid
phages.";
RL New Biol. 3:297-308(1991).
[2]
RP SEQUENCE OF 1-64 FROM N.A.
RC SPECIES=Phage 434;
RX MEDLINE=88167849; PubMed=2965063;
RA Limberger R.J., Campbell A.M.;
RT "Functional elements of DNA upstream from the integrase operon that

```

```

RT are conserved in bacteriophages 434 and lambda.";
RL Gene 61:135-144(1987).
[3]
RP SEQUENCE FROM N.A.
RC SPECIES=Phage HK022;
RX MEDLINE=89342457; PubMed=2547971;
RA Yagil E., Dolev S., Oberto J., Kisilev N., Ramaiiah N., Weisberg R.A.;
RT "Determinants of site-specific recombination in the lambdaoid
coliphage HK022. An evolutionary change in specificity.";
J. Mol. Biol. 207:695-717(1989)
[1] FUNCTION: EXCISIONASE AND INTEGRASE ARE NECESSARY FOR THE EXCISION
OF PROPHAGE FROM THE HOST GENOME BY SITE-SPECIFIC RECOMBINATION AT
THE ATT SITE.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.ebi.ac.uk/announce/
CC or send an email to license@ebi.ac.uk).
-----
DR EMBL; M60848; AAA67901.1; -
DR EMBL; X51962; CAA36222.1; -
DR PIR; S06533; S06533.
KW DNA recombination; DNA-binding.
SQ SEQUENCE 72 AA; 8635 MW; 0E6A484350344AA CRC64;

Query Match
Best Local Similarity 100.0%; Score 6; DB 1; Length 72;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 34 ARORRP 39
DB 10 ARORRP 15

RESULT 12
VXIS_LAMB STANDARD; PRT; 72 AA.
ID VXIS_LAMB
AC P03699;
DT 21-JUL-1986 (Rel. 01, Created)
DT 21-JUL-1986 (Rel. 01, Last sequence update)
DT 01-AUG-1992 (Rel. 23, Last annotation update)
DE Excisionase.
GN XIS.
OS Bacteriophage lambda.
OC Viruses; dsDNA viruses, no RNA stage; Caudovirales; Siphoviridae;
OC Lambda-like viruses.
OX NCBI_TaxID=10710;
[1]
RP SEQUENCE FROM N.A.
RX MEDLINE=63189071; PubMed=6221115;
RA Sanger F., Coulson A.R., Hong G.F., Hill D.F., Petersen G.B.;
RT "Nucleotide sequence of bacteriophage lambda DNA.";
J. Mol. Biol. 162:729-773(1982).
[2]
RP SEQUENCE FROM N.A.
RX MEDLINE=81053845; PubMed=6253947;
RA Davies R.W.;
RT "DNA sequence of the int-xis-pi region of the bacteriophage lambda;
RT overlap of the int and xis genes.";
Nucleic Acids Res. 8:1765-1782(1980).
[3]
RP SEQUENCE FROM N.A.
RX MEDLINE=80234646; PubMed=6446713;
RA Hoess R.H., Foeller C., Bidwell K., Landy A.;
RT "Site-specific recombination functions of bacteriophage lambda: DNA
RT sequence of regulatory regions and overlapping structural genes for
RT int and xis.";
Proc. Natl. Acad. Sci. U.S.A. 77:2482-2486(1980).
[1] FUNCTION: EXCISIONASE AND INTEGRASE ARE NECESSARY FOR THE EXCISION
OF PROPHAGE FROM THE HOST GENOME BY SITE-SPECIFIC RECOMBINATION AT

```

```
CC THE ATT SITE.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: J02459; AAA96563.1; -.
DR PIR: A04321; RSBPXL.
DR DNA recombination; DNA-binding.
SQ SEQUENCE 72 AA; 8605 MW; 0E6A4843502200AA CRC64;

Query Match 5.7%; Score 6; DB 1; Length 72;
Best Local Similarity 100.0%; Pred. No. 29;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 34 ARORRP 39
   |||||
Db 10 ARORRP 15

RESULT 13
Y10K_MDV STANDARD; PRT; 90 AA.
AC P06849;
DT 01-JAN-1988 (Rel. 06, Created)
DT 01-JAN-1988 (Rel. 06, Last sequence update)
DT 01-APR-1990 (Rel. 14, Last annotation update)
DE Hypothetical 10 kDa protein.
OS Wheat dwarf virus (WDV).
OC Viruses; ssDNA viruses; Geminiviridae; Mastrevirus.
OX NCBI_TaxID=10834;
RN [1]
RP SEQUENCE FROM N.A.
RA McDowell S.W., McDonald H., Hamilton W.D.O., Coutts R.H.A.,
RA Buck K.W.;
RT "The nucleotide sequence of cloned wheat dwarf virus DNA.";
RL EMO J. 4:2173-2180(1985).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: X02869; CAA26621.1; -.
DR PIR: D24356; D24356.
DR InterPro: IPR002621; Gemini_mov.
DR Pfam: PF01708; Gemini_mov; I.
RW Hypothetical protein.
SQ SEQUENCE 90 AA; 10146 MW; 0C5C46F0BCCD60E CRC64;

Query Match 5.7%; Score 6; DB 1; Length 90;
Best Local Similarity 100.0%; Pred. No. 35;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 8 NGDQOG 13
   |||||
Db 82 NGDQOG 87

RESULT 14
Y123_ECOLI STANDARD; PRT; 97 AA.
AC P19778; P76918;
DT 01-FEB-1991 (Rel. 17, Created)
DT 01-FEB-1991 (Rel. 17, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
```

```
DE Insertion element IS2 hypothetical 11.1 kDa protein (ORF3).
OS Escherichia coli.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
OX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=K12;
RX MEDLINE=88137965; PubMed=2830172;
RA Ronecker H.J., Rak B.;
RT "Genetic organization of insertion element IS2 based on a revised
RT nucleotide sequence.";
RL Gene 59:291-296(1987).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=K12;
RX MEDLINE=97251357; PubMed=9097039;
RA Alba H., Baba T., Fujita K., Hayashi K., Inada T., Isono K.,
RA Itoh T., Kasai H., Kashimoto K., Kimura S., Kitakawa M., Mori T.,
RA Kitagawa M., Makino K., Miki T., Mizobuchi K., Mori H., Mori T.,
RA Motomura K., Nakade S., Nakamura Y., Nishimoto H., Nishio Y.,
RA Oshima T., Saito N., Sempel G., Seki Y., Sivasundaram S.,
RA Tagami H., Takeda J., Takemoto K., Takeuchi Y., Wada C.,
RA Yamamoto Y., Horiuchi T.;
RT "A 570-Kb DNA sequence of the Escherichia coli K-12 genome
RT corresponding to the 28.0-40.1 min region on the linkage map.";
RL DNA Res. 3:363-377(1996).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: V00279; CAA23540.1; -.
DR EMBL: D90778; BAA15012.1; -.
DR EMBL: D90779; BAA15018.1; -.
DR EMBL: D90838; BAA15821.1; -.
DR EMBL: D90850; BAA16004.1; -.
DR EMBL: D90851; BAA16012.1; -.
DR EMBL: D90852; BAA22516.1; ALT_INIT.
DR PIR: J00041; J00041.
RW Transposable element; Hypothetical protein.
KV Transposable element; Hypothetical protein.
SQ SEQUENCE 97 AA; 11147 MW; DBB586E0BBA6B CRC64;

Query Match 5.7%; Score 6; DB 1; Length 97;
Best Local Similarity 100.0%; Pred. No. 38;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 84 IVGYGR 89
   |||||
Db 44 IVGYGR 49

RESULT 15
CMGC_BACHD STANDARD; PRT; 102 AA.
ID CMGC_BACHD
AC Q96923;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Comg operon protein 3 homolog precursor.
GN COMGC OR BH2827.
OS Bacillus halodurans.
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OX NCBI_TaxID=86665;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C-125 / JCM 9153;
RX MEDLINE=20512582; PubMed=11058132;
RA Takami H., Nakaone K., Takaki Y., Maeno G., Sasaki R., Masui N.,
```

```

RA Fuji F., Hirama C., Nakamura Y., Ogasawara N., Kuhara S.,
RA Horikoshi K.;
RT "Complete genome sequence of the alkaliphilic bacterium Bacillus
RT halodurans and genomic sequence comparison with Bacillus subtilis.";
RL Nucleic Acids Res. 28:4317-4331(2000).
CC
CC -1- FUNCTION: REQUIRED FOR TRANSFORMATION AND DNA-BINDING (BY
CC SIMILARITY).
CC -1- SUBUNIT: HOMODIMER (BY SIMILARITY).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AP001516; BAB06546.1; -.
DR InterPro; IPR000983; Bac_GSPG.
DR InterPro; IPR002416; Bac_GSPH.
DR InterPro; IPR001120; Prok_N_methyltn.
DR PRINTS; PR00813; BCTERIALGSG.
DR PRINTS; PR00885; BCTERIALGSG.
DR PROSITE; PS00409; PROKAR_NTER_METHYL_1.
KW Transport; Methylation; Transmembrane; Complete proteome.
FT PROPEP 1 10
FT CHAIN 1 102
FT TRANSMEM 11 31
FT MOD RES 11 11
FT DISULFID 46 85
FT FT 102 AA; 11368 MW; 3C4BD89B08564A43 CRC64;
SQ
SEQUENCE
Query Match 5.7%; Score 6; DB 1; Length 102;
Best Local Similarity 100.0%; Pred. No. 39;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 4 LTRNG 9
Db 91 LTRNG 96
RESULT 16
ID YN48 ARCFU STANDARD; PRT; 112 AA.
AC 030321;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical protein AF2348.
GN AF2348.
OS Archaeoglobus fulgidus.
OC Archaea; Euryarchaeota; Archaeoglobi; Archaeoglobales;
OC Archaeoglobaceae; Archaeoglobus.
OX NCBI_TaxID=2234;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=VC-16 / DSM 4304 / ATCC 49558;
RX MEDLINE=98049343; PubMed=9389475;
RA Kleink H.-P., Clayton R.A., Tomb J.-F., White O., Nelson K.E.,
RA Ketchum K.A., Dodson R.J., Gwinn M., Hickey E.K., Peterson J.D.,
RA Richardson D.L., Kerlavage A.R., Graham D.E., Kyriades N.C.,
RA Fleischmann R.D., Quackenbush J., Lee N.H., Sutton G.G., Gill S.,
RA Kirness E.F., Dougherty B.A., McKenney K., Adams M.D., Loftus B.,
RA Peterson S., Reich C.I., McNeil L.K., Badger J.H., Glodek A., Zhou L.,
RA Overbeek R., Goodyear U.D., Weidman J.F., McDonald L., Uterback T.,
RA Cotton M.D., Spriggs T., Artlich P., Kaine B.P., Sykes S.M.,
RA Sadow T.W., D'Andrea K.P., Bowman C., Fujii C., Garland S.A.,
RA Mason T.M., Olsen G.J., Fraser C.M., Smith H.O., Woese C.R.,
RA Venter J.C.;
RT "The complete genome sequence of the hyperthermophilic, sulphate-
RT reducing archaeon Archaeoglobus fulgidus.";
RL Nature 390:364-370(1997).

```

```

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AF001114; AAB91315.1; -.
DR TIGR; AF2348; -.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 112 AA; 13012 MW; C9C6AC0XDD6AC730 CRC64;
Query Match 5.7%; Score 6; DB 1; Length 112;
Best Local Similarity 100.0%; Pred. No. 43;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 95 DPSRKN 100
Db 80 DPSRKN 85
RESULT 17
ID SR14 CAEBL STANDARD; PRT; 116 AA.
AC 016927;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE Signal recognition particle 14 kDa protein homolog (SRP14).
GN P2566.8.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditioidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=Bristol N2;
RA Nelson J., Wohlmann P.;
RL Submitted (SRP-1997) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: SIGNAL-RECOGNITION-PARTICLE ASSEMBLY HAS A CRUCIAL ROLE
CC IN TARGETING SECRETORY PROTEINS TO THE ROUGH ENDOPLASMIC RETICULUM
CC MEMBRANE. SRP9 TOGETHER WITH SRP14 AND THE AU PORTION OF THE SRP
CC RNA, CONSTITUTES THE ELONGATION ARREST DOMAIN OF SRP. THE COMPLEX
CC OF SRP9 AND SRP14 IS REQUIRED FOR SRP RNA BINDING (BY SIMILARITY).
CC -1- SUBUNIT: SIGNAL RECOGNITION PARTICLE CONSISTS OF A 7S RNA MOLECULE
CC OR 300 NUCLEOTIDES AND SIX PROTEIN SUBUNITS: SRP72, SRP68, SRP54,
CC SRP19, SRP14 AND SRP9 (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: Cytoplasmic.
CC -1- SIMILARITY: BELONGS TO THE SRP14 FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AF022973; AAC25801.1; -.
DR HSSP; P16254; 1914.
DR WormPep; F2566.8; CE09644.
DR InterPro; IPR000210; SRP14.
DR Pfam; PF02290; SRP14; 1.
KW Signal recognition particle; RNA-binding.
SQ SEQUENCE 116 AA; 13203 MW; AE7CFEC80A1AA4C5 CRC64;
Query Match 5.7%; Score 6; DB 1; Length 116;
Best Local Similarity 100.0%; Pred. No. 44;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 55 AQWVNL 60

```

Db 93 AOMVNL 98

RESULT 18

YMG3\_ARCFU STANDARD: PRT: 122 AA.

AC 028021;

DT 16-OCT-2001 (Rel. 40, Created)

DT 16-OCT-2001 (Rel. 40, Last sequence update)

DT 16-OCT-2001 (Rel. 40, Last annotation update)

DE Hypothetical protein AF2263.

GN AF2263.

OS Archaeoglobus fulgidus.

OC Archaea; Euryarchaeota; Archaeoglobi; Archaeoglobales;

OC Archaeoglobaceae; Archaeoglobus.

OX NCBI\_TaxID=2234;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=VC-16 / DSM 4304 / ATCC 49558;

RX MEDLINE=98049343; PubMed=9389475;

RA Kleink H.-P., Clayton R.A., Tomb J.-F., White O., Nelson K.E., Ketchum K.A., Dodson R.J., Gwinn M., Hickey E.K., Peterson J.D., Richardson D.L., Kerevanage A.R., Graham D.E., Kyriides N.C., Fleischmann R.D., Quackenbush J., Lee N.H., Sutton G.G., Gill S., Kitkess B.F., Dougherty B.A., McKenney K., Adams M.D., Loftus B., Peterson S., Reich C.I., McNeil L.K., Badger J.H., Glodek A., Zhou L., Overbeek R., Gocayne J.D., Weidman J.F., McDonald L., Utecherack T., Cotton M.D., Spriggs T., Artlich P., Kaine B.P., Sykes S.M., Sadow P.W., D'Andrea K.P., Bowman C., Fujii C., Garland S.A., Macon T.M., Olsen G.J., Fraser C.M., Smith H.O., Woese C.R., R.A. Venter J.C.;

RT "The complete genome sequence of the hyperthermophilic, sulphate-reducing archaeon Archaeoglobus fulgidus.";

RL Nature 390:364-370(1997).

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC -----

CC EMBL; AE000948; AAB88987.1; -

DR TIGR; AF2263; -

KW Hypothetical protein: Coiled coil; Complete proteome.

FT DOMAIN 79 114 COILED COIL (POTENTIAL).

SEQUENCE 122 AA; 13824 MW; 31D01F2C8B0AA370 CRC64;

Query Match 5.7%; Score 6; DB 1; Length 122;

Best Local Similarity 100.0%; Pred. No. 46;

Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 86 GYGRDD 91

Db 28 GYGRDD 33

RESULT 19

RS15\_HUMAN STANDARD: PRT: 144 AA.

ID RS15\_HUMAN

AC P11174;

DT 01-JUL-1989 (Rel. 11, Created)

DT 01-AUG-1991 (Rel. 19, Last sequence update)

DT 15-JUN-2002 (Rel. 41, Last annotation update)

DE 40S ribosomal protein S15 (R1G protein).

GN RPS15 OR R1G.

OS Homo sapiens (Human);

OS Mus musculus (Mouse);

OS Rattus norvegicus (Rat);

OS Mesocricetus auratus (Golden hamster);

OS Sus scrofa (Pig); and

OS Gallus gallus (Chicken).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

OC NCBI\_TaxID=9606, 10090, 10116, 10036, 9823, 9031;

RN [1]

RP SEQUENCE FROM N.A.

RC SPECIES=Human; and M. auratus;

RX MEDLINE=88016150; PubMed=2821540;

RA Inoue C., Shiga K., Takasawa S., Kitagawa M., Yamamoto H., Okamoto H.;

RT "Evolutionary conservation of the insulinoma gene rig and its possible function.";

RL Proc. Natl. Acad. Sci. U.S.A. 84:6659-6662(1987).

RN [2]

RP SEQUENCE FROM N.A.

RC SPECIES=Human;

RX MEDLINE=90239060; PubMed=2159154;

RA Shiga K., Yamamoto H., Okamoto H.;

RT "Isolation and characterization of the human homologue of rig and its pseudogenes: the functional gene has features characteristic of housekeeping genes.";

RL Proc. Natl. Acad. Sci. U.S.A. 87:3594-3598(1990).

RN [3]

RP SEQUENCE FROM N.A.

RC SPECIES=Human; TISSUE=Brain;

RA Strauberg R.;

RL Submitted (NOV-2000) to the EMBL/GenBank/DBJ databases.

RN [4]

RP SEQUENCE FROM N.A.

RC SPECIES=Rat;

RX MEDLINE=87005616; PubMed=3019805;

RA Takasawa S., Yamamoto H., Terazono K., Okamoto H.;

RT "Novel gene activated in rat insulinomas.";

RL Diabetes 35:1178-1180(1986).

RN [5]

RP SEQUENCE FROM N.A.

RC SPECIES=Rat;

RX MEDLINE=9300333; PubMed=1390896;

RA Takasawa S., Tohgo A., Uno M., Shiga K., Yonekura H., Okamoto H.;

RT "The primary structure of rat rig/ribosomal protein S15 gene.";

RL Biochim. Biophys. Acta 1132:228-230(1992).

RN [6]

RP SEQUENCE.

RC SPECIES=Rat; STRAIN=Wistar; TISSUE=Liver;

RX MEDLINE=91257304; PubMed=2044758;

RA Kitagawa M., Takasawa S., Kikuchi N., Itoh T., Teraoka H., Yamamoto H., Okamoto H.;

RT "Rig encodes ribosomal protein S15. The primary structure of mammalian ribosomal protein S15.";

RL FEBS Lett. 283:210-214(1991).

RN [7]

RP SEQUENCE FROM N.A.

RC SPECIES=Mouse; and Chicken;

RX MEDLINE=90165964; PubMed=2154983;

RA Sugawara A., Nata K., Inoue C., Takasawa S., Yamamoto H., Okamoto H.;

RT "Nucleotide sequence determination of mouse, chicken and Xenopus laevis rig cDNAs: the rig-encoded protein is extremely conserved during vertebrate evolution.";

RL Biochem. Biophys. Res. Commun. 166:1501-1507(1990).

RN [8]

RP SEQUENCE FROM N.A.

RC SPECIES=Pig; TISSUE=Fat;

RA Nishimura M., Satoh H., Suzuki H., Hamasima N.;

RT "Cloning of the pig gene similar to the human insulinoma rig-analog RNA-binding protein.";

RL Submitted (JUL-1997) to the EMBL/GenBank/DBJ databases.

RN [9]

RP SEQUENCE FROM N.A.

RC SPECIES=Chicken;

RX MEDLINE=92084155; PubMed=1748316;

RA Sugawara A., Shiga K., Takasawa S., Yonekura H., Yamamoto H., Okamoto H.;

RT "Sequence of the chicken rig gene encoding ribosomal protein S15.";

RL Gene 108:313-314(1991).

```

CC -1- SIMILARITY: BELONGS TO THE S19P FAMILY OF RIBOSOMAL PROTEINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; J02984; AAA36036.1; -
DR EMBL; M32405; AAA36568.1; -
DR EMBL; BC000085; AAA00085.1; -
DR EMBL; M19393; AAA42044.1; -
DR EMBL; D11388; BAA01984.1; -
DR EMBL; J02983; AAA37094.1; -
DR EMBL; M33330; AAA40055.1; -
DR EMBL; M33331; AAA49057.1; -
DR EMBL; AB005624; BAA21510.1; -
DR EMBL; D10167; BAA01036.1; -
DR PIR; A35908; R3HUI5.
DR PIR; S07438; R3HUI5.
DR PIR; B33885; R3HUI5.
DR PIR; A34823; A34823.
DR PIR; B34823; B34823.
DR PIR; JH0510; JH0510.
DR PIR; S28939; S28939.
DR HSSP; P80381; 10KF.
DR Genew; HGNC:10386; RPS15.
DR MIM; 180535; -
DR MGD; MGI:98117; Rps15.
DR InterPro; IPR002222; Ribosomal_S19.
DR Pfam; PF00203; Ribosomal_S19; 1.
DR PRINTS; PR00975; RIBOSOMALS19.
DR ProDom; PD001012; Ribosomal_S19; 1.
DR TIGRFAMs; TIGR01025; rps5_arch; 1.
DR PROSITE; PS00323; RIBOSOMAL_S19; 1.
KW Ribosomal protein; Acetylation.
FT INIT MET 0
FT MOD RES 0
SQ SEQUENCE 144 AA; 16909 MW; D049BB9232C7CD36 CRC64;

Query Match 5.7%; Score 6; DB 1; Length 144;
Best Local Similarity 100.0%; Pred. No. 52;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 33 SARQR 38
Db 37 SARQR 42

RESULT 20
ID Y881_VIBCH STANDARD; PRT; 149 AA.
AC O9KTM1;
DT 15-JUN-2002 (Rel. 41, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Hypothetical protein VC0881.
GN VC0881.
OS Vibrio cholerae.
OC Bacteria; Proteobacteria; gamma subdivision; Vibrionaceae; Vibrio.
OX NCBI_TaxID=666;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=El Tor N16961 / Serotype O1;
RX MEDLINE=20406833; PubMed=10952301;
RA Heidelberg J.F., Elesen J.A., Nelson W.C., Clayton R.A., Gwin M.L.,
RA Dodson R.J., Helt D.H., Hickey E.K., Peterson J.D., Umayam L.A.,
RA Gill S.R., Nelson K.E., Reed T.D., Jettelin H., Richardson D.,
RA Ermoлева M.D., Vanachavan J., Bass S., Qin H., Dragol I., Sellers P.,
RA McDonald L., Uterback T., Fleischmann R.D., Nierman W.C., White O.,
RA Salzberg S.L., Smith H.O., Colwell R.R., Mekalanos J.J., Venter J.C.,

```

```

RA Fraser C.M.;
RT "DNA sequence of both chromosomes of the cholera pathogen Vibrio
RL Nature 406:477-483(2000).
CC -1- SIMILARITY: BELONGS TO THE UPF0178 FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AE004172; AAF94043.1; -
DR TIGR; VC0881; -
DR InterPro; IPR003791; DUF188.
DR Pfam; PF02639; DUF188; 1.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 149 AA; 16454 MW; F5071B3C16105FAD CRC64;

Query Match 5.7%; Score 6; DB 1; Length 149;
Best Local Similarity 100.0%; Pred. No. 54;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 45 SGFDIA 50
Db 52 SGFDIA 57

RESULT 21
ID MYG_AOTTR STANDARD; PRT; 153 AA.
AC P02151;
DT 21-JUL-1986 (Rel. 01, Created)
DT 21-JUL-1986 (Rel. 01, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Myoglobin.
DE Myoglobin.
GN MB.
OS Aotus trivirgatus (Night monkey) (Douroucoul).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Platyrrhini; Cebidae; Aotinae; Aotus.
OX NCBI_TaxID=9505;
RN [1]
RP SEQUENCE.
RC TISSUE=Skeletal muscle;
RX MEDLINE=84108869; PubMed=6692912;
RA Heindokel N., Lehmann H.;
RT "The myoglobin of primates: the Night Monkey, Aotus trivirgatus
RL Cebidae, Platyrrhini, Anthropoidea)." ;
CC FBBS Lett. 165:46-50(1984).
CC -1- FUNCTION: SERVES AS A RESERVE SUPPLY OF OXYGEN AND FACILITATES
CC THE MOVEMENT OF OXYGEN WITHIN MUSCLES.
CC -1- SIMILARITY: BELONGS TO THE GLOBIN FAMILY.
DR PIR; A02471; MYMON.
DR HSSP; P02189; 1MWC.
DR InterPro; IPR000971; Globin.
DR InterPro; IPR002335; Myoglobin.
DR Pfam; PF00042; globin; 1.
DR PRINTS; PR00613; MYOGLOBIN.
DR PROSITE; PS01033; GLOBIN; 1.
KW Heme; Oxygen transport; Transport; Muscle.
FT METAL 64
FT METAL 64 IRON (HEME PROXIMAL LIGAND).
FT METAL 93 IRON (HEME DISTAL LIGAND).
SQ SEQUENCE 153 AA; 17030 MW; E8EDB8D461D020E CRC64;

Query Match 5.7%; Score 6; DB 1; Length 153;
Best Local Similarity 100.0%; Pred. No. 55;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 101 GGIILKK 106
Db 73 GGIILKK 78

```

```
RESULT 22
MYG_BALAC STANDARD; PRT; 153 AA.
ID MYG_BALAC
AC P02179;
DT 21-JUL-1986 (Rel. 01, Created)
DT 21-JUL-1986 (Rel. 01, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Myoglobin.
GN MB.
OS Balaenoptera acutorostrata (Minke whale) (lesser rorqual).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Cetacea; Mysticeti;
OC Balaenopteridae; Balaenoptera.
OX NCBI_TaxId=9767;
RN [1]
RP SEQUENCE.
RC TISSUE=Skeletal muscle;
RX MEDLINE=7112438; PubMed=836810;
RA Lehman L.D., Dwyer F.E., Bogardt R.A. Jr., Jones B.N., Gurd F.R.N.;
RT "The complete amino acid sequence of the major component myoglobin
from the arctic minke whale, Balaenoptera acutorostrata.";
RL Biochemistry 16:706-709(1977).
CC -1- FUNCTION: SERVES AS A RESERVE SUPPLY OF OXYGEN AND FACILITATES
THE MOVEMENT OF OXYGEN WITHIN MUSCLES.
CC -1- SIMILARITY: BELONGS TO THE GLOBIN FAMILY.
DR HSSP; P02185; 1BZ6.
DR InterPro; IPR000971; Globin.
DR Pfam; PF00042; Globin; 1.
DR PRINTS; PR00613; MYOGLOBIN.
DR PROSITE; PS01033; GLOBIN; 1.
KW Heme; Oxygen transport; Transport; Muscle.
FT METAL 64 64 IRON (HEME DISTAL LIGAND).
FT METAL 93 93 IRON (HEME PROXIMAL LIGAND).
SQ SEQUENCE 153 AA; 17155 MW; A5364871B9705C6E CRC64;

Query Match
Best local similarity 100.0%; Score 6; DB 1; Length 153;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 101 GGLTKK 106
Db 73 GGLTKK 78

RESULT 23
MYG_BALPH STANDARD; PRT; 153 AA.
ID MYG_BALPH
AC P02180;
DT 21-JUL-1986 (Rel. 01, Created)
DT 21-JUL-1986 (Rel. 01, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Myoglobin.
GN MB.
OS Balaenoptera physalus (Finback whale) (Common rorqual).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Cetacea; Mysticeti;
OC Balaenopteridae; Balaenoptera.
OX NCBI_TaxId=9770;
RN [1]
RP SEQUENCE.
RC TISSUE=Skeletal muscle;
RX MEDLINE=7818712; PubMed=656375;
RA Dimarichi R.D., Wang C.-C., Hemenway J.B., Gurd F.R.N.;
RT "Complete amino acid sequence of the major component myoglobin of
finback whale (Balaenoptera physalus).";
RL Biochemistry 17:1968-1970(1978).
CC -1- FUNCTION: SERVES AS A RESERVE SUPPLY OF OXYGEN AND FACILITATES
THE MOVEMENT OF OXYGEN WITHIN MUSCLES.
CC -1- SIMILARITY: BELONGS TO THE GLOBIN FAMILY.
```

```
DR PIR; A02502; MYWHF.
DR HSSP; P02185; 1BZ6.
DR InterPro; IPR000971; Globin.
DR InterPro; IPR002335; Myoglobin.
DR Pfam; PF00042; Globin; 1.
DR PRINTS; PR00613; MYOGLOBIN.
DR PROSITE; PS01033; GLOBIN; 1.
KW Heme; Oxygen transport; Transport; Muscle.
FT METAL 64 64 IRON (HEME DISTAL LIGAND).
FT METAL 93 93 IRON (HEME PROXIMAL LIGAND).
SQ SEQUENCE 153 AA; 17086 MW; 8AFC89C1BF00291F CRC64;

Query Match
Best local similarity 100.0%; Score 6; DB 1; Length 153;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 101 GGLTKK 106
Db 73 GGLTKK 78

RESULT 24
MYG_BOVIN STANDARD; PRT; 153 AA.
ID MYG_BOVIN
AC P02192;
DT 21-JUL-1986 (Rel. 01, Created)
DT 01-AUG-1990 (Rel. 15, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE Myoglobin.
GN MB.
OS Bos taurus (Bovine).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
OC Bovidae; Bovinae; Bos.
OX NCBI_TaxId=9913;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=89278066; PubMed=2659586;
RA Shimada H., Fukasawa T., Ishimura Y.;
RT "Expression of bovine myoglobin cDNA as a functionally active
holoprotein in Saccharomyces cerevisiae.";
RL J. Biochem. 105:417-422(1989).
RN [2]
RP SEQUENCE.
RX MEDLINE=71028010; PubMed=5477290;
RA Han X., Dautrevaux M., Chaila X., Biserte G.;
RT "The covalent structure of beef heart myoglobin.";
RL Eur. J. Biochem. 16:465-471(1970).
CC -1- FUNCTION: SERVES AS A RESERVE SUPPLY OF OXYGEN AND FACILITATES
THE MOVEMENT OF OXYGEN WITHIN MUSCLES.
CC -1- SIMILARITY: BELONGS TO THE GLOBIN FAMILY.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
-----
CC EMBL; D00409; BAA00311.1; -.
DR PIR; A02513; MYBO.
DR PIR; JX0068; JX0068.
DR HSSP; P02188; 1HRM.
DR InterPro; IPR000971; Globin.
DR InterPro; IPR002335; Myoglobin.
DR Pfam; PF00042; Globin; 1.
DR PRINTS; PR00613; MYOGLOBIN.
DR PROSITE; PS01033; GLOBIN; 1.
KW Heme; Oxygen transport; Transport; Muscle.
FT INIT_MET 0 0
FT METAL 64 64 IRON (HEME DISTAL LIGAND).
FT METAL 93 93 IRON (HEME PROXIMAL LIGAND).
```



FT CONFLICT 9 9 L -> A (IN REF. 2).  
 FT CONFLICT 99 101 IPV -> VIP (IN REF. 2).  
 FT CONFLICT 122 124 DFG -> NFA (IN REF. 2).  
 FT CONFLICT 142 145 MAAO -> AAEK (IN REF. 2).  
 SQ SEQUENCE 153 AA; 16946 MW; 4E69AC6BE2101728 CRC64;

Query Match  
 Best Local Similarity 5.7%; Score 6; DB 1; Length 153;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 101 GGLLKK 106  
 Db 73 GGLLKK 78

RESULT 25  
 MYG\_CALJA STANDARD; PRT; 153 AA.

AC P02152;  
 DT 21-JUL-1986 (Rel. 01, Created)  
 DT 21-JUL-1986 (Rel. 01, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Myoglobin.

OS Callitrichus jacchus (Common marmoset).  
 OC Eukaryota; Metazoa; Chordata; Cranialata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Platyrrhini; Callitrichidae;  
 OC Callitrichus.  
 OC NCBI\_Taxid=9483;  
 RN [1]  
 RP SEQUENCE.

RC TISSUE=Skeletal muscle;  
 RX MEDLINE=73231297; PubMed=4198766;  
 RA Romero-Herrera A.E., Lehmann H.;  
 RT "The myoglobin of primates. 4. New World monkeys: Cebidae: (1) *Leontideus rosalia* (squirrel monkey); (2) *Lagothrix lagotricha* (Humboldt's woolly monkey). Callitrichidae: Callitrichus jacchus (common marmoset).";  
 RL Biochim. Biophys. Acta 317:65-84(1973).  
 RN [2]

RP PARTIAL SEQUENCE (MINOR COMPONENT).  
 RA Romero-Herrera A.E., Lehmann H.;  
 RT "N-terminal chain elongation as evidence for duplication of myoglobin in three South American monkeys.";  
 RL FEBS Lett. 31:175-180(1973).  
 CC -1- FUNCTION: SERVES AS A RESERVE SUPPLY OF OXYGEN AND FACILITATES THE MOVEMENT OF OXYGEN WITHIN MUSCLES.  
 CC -1- MISCELLANEOUS: MARMOSET, WOOLLY MONKEY, AND SQUIRREL MONKEY HAVE A MINOR MYOGLOBIN COMPONENT THAT APPEARS TO DIFFER FROM EACH MAJOR COMPONENT IN HAVING PHE-LEYS PRECEDING POSITION 1.  
 CC -1- SIMILARITY: BELONGS TO THE GLOBIN FAMILY.  
 DR HSP; P02189; IMWC.  
 DR InterPro; IPR000971; Globin.  
 DR InterPro; IPR002335; Myoglobin.  
 DR Pfam; PF00042; Globin; 1.  
 DR PRINTS; PR00613; MYOGLOBIN.  
 DR PROSITE; PS01033; GLOBIN; 1.

KW Heme; Oxygen transport; Transport; Muscle.  
 FT METAL 64 64 IRON (HEME DISTAL LIGAND).  
 FT METAL 93 93 IRON (HEME PROXIMAL LIGAND).  
 FT VARIANT 1 1 G -> FRG (IN MINOR COMPONENT).  
 FT VARIANT 62 62 K -> R.  
 SQ SEQUENCE 153 AA; 17044 MW; BEBDA0776B5F109B CRC64;

Query Match  
 Best Local Similarity 5.7%; Score 6; DB 1; Length 153;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 101 GGLLKK 106  
 Db 73 GGLLKK 78

RESULT 26  
 MYG\_CANFA STANDARD; PRT; 153 AA.

AC P02158;  
 DT 21-JUL-1986 (Rel. 01, Created)  
 DT 21-JUL-1986 (Rel. 01, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Myoglobin.

OS Canis familiaris (Dog), and  
 OS Otocyon megalotis (Bat-eared fox).  
 OC Eukaryota; Metazoa; Chordata; Cranialata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.  
 OC NCBI\_Taxid=9615, 9624;  
 RN [1]  
 RP SEQUENCE.

RC SPECIES=C.familiaris;  
 RX MEDLINE=76136443; PubMed=1252463;  
 RA Dumur V., Dautrevaux M., Han K.;  
 RT "The covalent structure of dog myoglobin.";  
 RL Biochim. Biophys. Acta 420:376-386(1976).

RN [2]  
 RP REVISIONS.  
 RC SPECIES=C.familiaris;  
 RA Dumur V., Dautrevaux M., Han K.;  
 RL Biochim. Biophys. Acta 427:759-761(1976).  
 RN [3]

RP SEQUENCE.  
 RC SPECIES=O.megalotis;  
 RX MEDLINE=77065657; PubMed=793619;  
 RA Darbre P.D., Lehmann H.;  
 RT "Comparison of the myoglobin of the bat-eared fox (*Otocyon megalotis*) with that of the domestic dog (*Canis familiaris*).";  
 RL Biochim. Biophys. Acta 453:285-288(1976).  
 CC -1- FUNCTION: SERVES AS A RESERVE SUPPLY OF OXYGEN AND FACILITATES THE MOVEMENT OF OXYGEN WITHIN MUSCLES.  
 CC -1- SIMILARITY: BELONGS TO THE GLOBIN FAMILY.  
 DR PIR; A02478; MYDG.  
 DR PIR; A02478; MYEXB.

DR HSP; P02189; IMWC.  
 DR HSC-2DPAGE; P02158; DOG.  
 DR InterPro; IPR000971; Globin.  
 DR InterPro; IPR002335; Myoglobin.  
 DR Pfam; PF00042; Globin; 1.  
 DR PRINTS; PR00613; MYOGLOBIN.  
 DR PROSITE; PS01033; GLOBIN; 1.  
 DR Heme; Oxygen transport; Transport; Muscle.  
 FT METAL 64 64 IRON (HEME DISTAL LIGAND).  
 FT METAL 93 93 IRON (HEME PROXIMAL LIGAND).  
 SQ SEQUENCE 153 AA; 17206 MW; 94C7C1B43EB7B0F1 CRC64;

Query Match  
 Best Local Similarity 5.7%; Score 6; DB 1; Length 153;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 101 GGLLKK 106  
 Db 73 GGLLKK 78

RESULT 27  
 MYG\_CEBAP STANDARD; PRT; 153 AA.

AC P02153;  
 DT 21-JUL-1986 (Rel. 01, Created)  
 DT 21-JUL-1986 (Rel. 01, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Myoglobin.

OS Cebus apella (Brown-capped capuchin).  
 OS Eukaryota; Metazoa; Chordata; Cranialata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Platyrrhini; Cebidae; Cebus.

OX NCBI\_TaxID=9515;  
 RN [1]  
 RP SEQUENCE.  
 RC TISSUE=skeletal muscle;  
 RX MEDLINE=81021734; PubMed=7417495;  
 RA Dene H., Sazy U., Romero-Herrera A.E.;  
 RT "The myoglobin of primates X."  
 RL Blochm. Biophys. Acta 625:133-145 (1980).  
 CC -1- FUNCTION: SERVES AS A RESERVE SUPPLY OF OXYGEN AND FACILITATES  
 THE MOVEMENT OF OXYGEN WITHIN MUSCLES.  
 CC -1- SIMILARITY: BELONGS TO THE GLOBIN FAMILY.  
 DR PIR; A02473; MYOC.  
 DR HSP; P02189; IMWC.  
 DR InterPro; IPR000971; Globin.  
 DR InterPro; IPR002335; Myoglobin.  
 DR Pfam; PF00042; globin; 1.  
 DR PRINTS; PR00613; MYOGLOBIN.  
 DR PROSITE; PS01033; GLOBIN; 1.  
 KW Heme; Oxygen transport; Transport; Muscle.  
 FT METAL 64 64 IRON (HEME DISTAL LIGAND).  
 FT METAL 93 93 IRON (HEME PROXIMAL LIGAND).  
 SQ SEQUENCE 153 AA; 17007 MW; 10D9EC05F041086 CRC64;

Query Match 5.7%; Score 6; DB 1; Length 153;  
 Best Local Similarity 100.0%; Pred. No. 55;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 101 GGLIKK 106  
 |||||  
 Db 73 GGLIKK 78

RESULT 28  
 MYG\_CEREL STANDARD; PRT; 153 AA.  
 AC P02191;  
 DT 21-JUL-1986 (Rel. 01, Created)  
 DT 21-JUL-1986 (Rel. 01, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Myoglobin.  
 GN MB.  
 OS Cervus elaphus (Red deer).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Cervoidae;  
 OC Cervidae; Cervinae; Cervus.  
 OX NCBI\_TaxID=9860;  
 RN [1]  
 RP SEQUENCE.  
 RC TISSUE=skeletal muscle;  
 RX MEDLINE=76072125; PubMed=1202228;  
 RA Vortsch W., Anderer F.A.;  
 RT "The amino acid sequence of myoglobin from skeletal muscles of red  
 deer (Cervus elaphus)."  
 RL J. Mol. Evol. 5:315-326 (1975).  
 CC -1- FUNCTION: SERVES AS A RESERVE SUPPLY OF OXYGEN AND FACILITATES  
 THE MOVEMENT OF OXYGEN WITHIN MUSCLES.  
 CC -1- SIMILARITY: BELONGS TO THE GLOBIN FAMILY.  
 DR PIR; A02512; MYDE.  
 DR HSP; P02188; IMWA.  
 DR InterPro; IPR000971; Globin.  
 DR InterPro; IPR002335; Myoglobin.  
 DR Pfam; PF00042; globin; 1.  
 DR PRINTS; PR00613; MYOGLOBIN.  
 DR PROSITE; PS01033; GLOBIN; 1.  
 KW Heme; Oxygen transport; Transport; Muscle.  
 FT METAL 64 64 IRON (HEME DISTAL LIGAND).  
 FT METAL 93 93 IRON (HEME PROXIMAL LIGAND).  
 SQ SEQUENCE 153 AA; 16922 MW; A6FE1A6BE2181722 CRC64;

Query Match 5.7%; Score 6; DB 1; Length 153;  
 Best Local Similarity 100.0%; Pred. No. 55;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 101 GGLIKK 106  
 |||||  
 Db 73 GGLIKK 78

RESULT 29  
 MYG\_ELEMA STANDARD; PRT; 153 AA.  
 AC P02186;  
 DT 21-JUL-1986 (Rel. 01, Created)  
 DT 01-NOV-1997 (Rel. 35, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Myoglobin.  
 GN MB.  
 OS Elephas maximus (Indian elephant).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Proboscidea; Elephantiidae; Elephas.  
 OX NCBI\_TaxID=9783;  
 RN [1]  
 RP SEQUENCE.  
 RC TISSUE=skeletal muscle;  
 RX MEDLINE=80145813; PubMed=6102395;  
 RA Dene H., Goodman M., Romero-Herrera A.E.;  
 RT "The amino acid sequence of elephant (Elephas maximus) myoglobin and  
 the phylogeny of Proboscidea."  
 RL Proc. R. Soc. Lond., B, Biol. Sci. 207:111-127 (1980).  
 RN [2]  
 RP X-RAY CRYSTALLOGRAPHY (1.78 ANGSTROMS) AND REVISIONS TO 27 AND 29.  
 RX MEDLINE=95385531; PubMed=7657658;  
 RA Blsig D.A., di Iorio E.E., Diederichs K., Winterhalter K.H.,  
 RA Plontek K.;  
 RT "Crystal structure of Asian elephant (Elephas maximus) cyano-  
 myoglobin at 1.78-A resolution. Phe29(B10) accounts for its  
 unusual ligand binding properties."  
 RT unusual ligand binding properties."  
 RL J. Biol. Chem. 270:20754-20762 (1995).  
 CC -1- FUNCTION: SERVES AS A RESERVE SUPPLY OF OXYGEN AND FACILITATES  
 THE MOVEMENT OF OXYGEN WITHIN MUSCLES.  
 CC -1- SIMILARITY: BELONGS TO THE GLOBIN FAMILY.  
 DR PIR; A02507; MYELI.  
 DR PDB; 1EWY; 20-APR-95.  
 DR InterPro; IPR000971; Globin.  
 DR InterPro; IPR002335; Myoglobin.  
 DR Pfam; PF00042; globin; 1.  
 DR PRINTS; PR00613; MYOGLOBIN.  
 DR PROSITE; PS01033; GLOBIN; 1.  
 KW Heme; Oxygen transport; Transport; Muscle; 3D-structure.  
 FT METAL 93 93 IRON (HEME PROXIMAL LIGAND).  
 FT CONFLICT 27 27 T -> F (IN REF. 1).  
 FT CONFLICT 29 29 F -> L (IN REF. 1).  
 SQ SEQUENCE 153 AA; 16995 MW; 5823AB85F706EA9C CRC64;

Query Match 5.7%; Score 6; DB 1; Length 153;  
 Best Local Similarity 100.0%; Pred. No. 55;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 101 GGLIKK 106  
 |||||  
 Db 73 GGLIKK 78

RESULT 30  
 MYG\_ERIEU STANDARD; PRT; 153 AA.  
 AC P02156;  
 DT 21-JUL-1986 (Rel. 01, Created)  
 DT 21-JUL-1986 (Rel. 01, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Myoglobin.  
 GN MB.  
 OS Erinaceus europaeus (Western European hedgehog).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Insectivora; Erinaceidae; Erinaceus.  
 OX NCBI\_TaxID=9365;

RN [1]  
RP SEQUENCE.  
RC TISSUE=Skeletal muscle;  
RX MEDLINE=75109311; PubMed=1167790;  
RA Romero-Herrera A.E.; Lehmann H.; Fakes W.;  
RT "The primary structure of the myoglobin of the insectivore Erinaceus  
europeus (common European hedgehog).";  
RL Biochim. Biophys. Acta 379:13-21(1975).  
CC -1- FUNCTION: SERVES AS A RESERVE SUPPLY OF OXYGEN AND FACILITATES  
THE MOVEMENT OF OXYGEN WITHIN MUSCLES.  
CC -1- SIMILARITY: BELONGS TO THE GLOBIN FAMILY.  
DR HSBP; A02476; MYHC.  
DR InterPro; IPR000971; Globin.  
DR InterPro; IPR002335; Myoglobin.  
DR Pfam; PF00042; Globin; 1.  
DR PRINTS; PR00613; MYOGLOBIN.  
DR PROSITE; PS01033; GLOBIN; 1.  
KW Heme; Oxygen transport; Transport; Muscle.  
FT METAL 64 64 IRON (HEME DISTAL LIGAND).  
FT METAL 93 93 IRON (HEME PROXIMAL LIGAND).  
SQ SEQUENCE 153 AA; 16983 MW; 1D54633CA68F2216 CRC64;

Query Match 5.7%; Score 6; DB 1; Length 153;  
Best Local Similarity 100.0%; Pred. No. 55;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 101 GILKK 106  
Db 73 GILKK 78

RESULT 31  
MYG\_ESCGI STANDARD; PRT; 153 AA.  
AC P02177;  
DT 21-JUL-1986 (Rel. 01, Created)  
DT 21-JUL-1986 (Rel. 01, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE Myoglobin.  
GN MB.  
OS Eschrichtius gibbosus (California gray whale) (Eschrichtius robustus).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Cetartiodactyla; Cetacea; Mysticeti;  
OC Eschrichtidae; Eschrichtius.  
OX NCBI\_TaxID=9764;  
RN [1]  
RP SEQUENCE.  
RC TISSUE=Skeletal muscle;  
RX MEDLINE=7623191; PubMed=938629;  
RA Bogardt R.A. Jr.; Dulaet F.B.; Lehman L.D.; Jones B.N.; Gurd F.R.N.;  
RT "Complete primary structure of the major component myoglobin of  
California gray whale (Eschrichtius gibbosus).";  
RL Biochemistry 15:2597-2602(1976).  
CC -1- FUNCTION: SERVES AS A RESERVE SUPPLY OF OXYGEN AND FACILITATES  
THE MOVEMENT OF OXYGEN WITHIN MUSCLES.  
CC -1- SIMILARITY: BELONGS TO THE GLOBIN FAMILY.  
DR HSBP; A02499; MYHC.  
DR InterPro; IPR000971; Globin.  
DR InterPro; IPR002335; Myoglobin.  
DR Pfam; PF00042; Globin; 1.  
DR PRINTS; PR00613; MYOGLOBIN.  
DR PROSITE; PS01033; GLOBIN; 1.  
KW Heme; Oxygen transport; Transport; Muscle.  
FT METAL 64 64 IRON (HEME DISTAL LIGAND).  
FT METAL 93 93 IRON (HEME PROXIMAL LIGAND).  
SQ SEQUENCE 153 AA; 17118 MW; 4A5587BA0E805C7D CRC64;

Query Match 5.7%; Score 6; DB 1; Length 153;  
Best Local Similarity 100.0%; Pred. No. 55;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 101 GILKK 106  
Db 73 GILKK 78

RESULT 32  
MYG\_GALCR STANDARD; PRT; 153 AA.  
AC P02168;  
DT 21-JUL-1986 (Rel. 01, Created)  
DT 21-JUL-1986 (Rel. 01, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE Myoglobin.  
GN MB.  
OS Galago crassicaudatus (Thick-tailed galago) (Otolemur crassicaudatus).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Strepsirrhini; Galagonidae; Otolemur.  
OX NCBI\_TaxID=9463;  
RN [1]  
RP PARTIAL SEQUENCE.  
RX MEDLINE=74008816; PubMed=4582928;  
RA Romero-Herrera A.E.; Lehmann H.;  
RT "The myoglobin of primates. V. Prosimians: Galago crassicaudatus  
(thick-tailed galago) and Lepilemur mustelinus ( sportive Lemur).";  
RL Biochim. Biophys. Acta 322:10-22(1973).  
CC -1- FUNCTION: SERVES AS A RESERVE SUPPLY OF OXYGEN AND FACILITATES  
THE MOVEMENT OF OXYGEN WITHIN MUSCLES.  
CC -1- MISCELLANEOUS: SOME AMIDES WERE ASSIGNED BY ELECTROPHORETIC  
MOBILITY OR BY HOMOLOGY.  
CC -1- SIMILARITY: BELONGS TO THE GLOBIN FAMILY.  
DR HSBP; A02490; MYG.  
DR HSBP; P02189; MYG.  
DR InterPro; IPR000971; Globin.  
DR InterPro; IPR002335; Myoglobin.  
DR Pfam; PF00042; Globin; 1.  
DR PRINTS; PR00613; MYOGLOBIN.  
DR PROSITE; PS01033; GLOBIN; 1.  
KW Heme; Oxygen transport; Transport; Muscle.  
FT METAL 64 64 IRON (HEME DISTAL LIGAND).  
FT METAL 93 93 IRON (HEME PROXIMAL LIGAND).  
SQ SEQUENCE 153 AA; 16970 MW; 0DB58E67ADC5733 CRC64;

Query Match 5.7%; Score 6; DB 1; Length 153;  
Best Local Similarity 100.0%; Pred. No. 55;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 101 GILKK 106  
Db 73 GILKK 78

RESULT 33  
MYG\_GORBE STANDARD; PRT; 153 AA.  
AC P02147;  
DT 21-JUL-1986 (Rel. 01, Created)  
DT 21-JUL-1986 (Rel. 01, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE Myoglobin.  
GN MB.  
OS Gorilla gorilla beringei (Mountain gorilla) (Highland gorilla).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Gorilla.  
OX NCBI\_TaxID=9594;  
RN [1]  
RP SEQUENCE.  
RC TISSUE=Skeletal muscle;  
RX MEDLINE=75205685; PubMed=1148222;  
RA Romero-Herrera A.E.; Lehmann H.; Fosse D.;  
RT "The myoglobin of primates. VIII. Gorilla gorilla beringei (eastern  
highland gorilla).";  
RL Biochim. Biophys. Acta 393:363-388(1975).  
CC -1- FUNCTION: SERVES AS A RESERVE SUPPLY OF OXYGEN AND FACILITATES

```

CC      THE MOVEMENT OF OXYGEN WITHIN MUSCLES.
CC      -1- SIMILARITY: BELONGS TO THE GLOBIN FAMILY.
CC      PIR; A02467; MYGO.
CC      HSSP; P02144; 2MWL.
CC      DR InterPro; IPR000971; Globin.
CC      DR InterPro; IPR002335; Myoglobin.
CC      DR Pfam; PF00042; globin; 1.
CC      DR PRINTS; PR00613; MYOGLOBIN.
CC      DR PROSITE; PS01033; GLOBIN; 1.
CC      KM Heme; Oxygen transport; Transport; Muscle.
CC      FT METAL 64 64 IRON (HEME DISTAL LIGAND).
CC      FT METAL 93 93 IRON (HEME PROXIMAL LIGAND).
CC      SQ SEQUENCE 153 AA; 17042 MW; 5F876DC59EBA20D5 CRC64;

Query Match 5.7%; Score 6; DB 1; Length 153;
Best Local Similarity 100.0%; Pred. No. 55;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 101 GILK 106
   |||||
Db 73 GILK 78

RESULT 34
MYG_HALGR STANDARD; PRT; 153 AA.
AC P02162;
DT 21-JUL-1986 (Rel. 01, Created)
DT 21-JUL-1986 (Rel. 01, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Myoglobin.
GN MB.
OS Halichoerus grypus (Gray seal), and
OS Phoca vitulina (Harbor seal).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Carnivora; Plimipedia; Phocidae; Halichoerus.
OC NCBI_TaxID=9711, 9720;
OX [1]
RN [1]
RP SEQUENCE FROM N.A.
RC SPECIES=H. grypus;
RX MEDLINE=83141779; PubMed=6828156;
RA Blanchetot A., Wilson V., Wood D., Jeffreys A.J.;
RT "The seal myoglobin gene: an unusually long globin gene.";
RL Nature 301:732-734(1983).
RN [2]
RP SEQUENCE.
RC SPECIES=P. vitulina;
RX MEDLINE=69177451; PubMed=5782005;
RA Bradshaw R.A., Gurd F.R.N.;
RT "Comparison of myoglobins from harbor seal, porpoise, and sperm
RT whale. V. The complete amino acid sequences of harbor seal and
RT porpoise myoglobins.";
RL J. Biol. Chem. 244:2167-2181(1969).
RN [3]
RP X-RAY CRYSTALLOGRAPHY (2.5 ANGSTROMS).
RC SPECIES=P. vitulina;
RX MEDLINE=79132579; PubMed=745243;
RA Scouloudi H., Baker E.N.;
RT "X-ray crystallographic studies of seal myoglobin. The molecule at
RT 2.5-A resolution.";
RL J. Mol. Biol. 126:637-660(1978).
CC      -1- FUNCTION: SERVES AS A RESERVE SUPPLY OF OXYGEN AND FACILITATES
CC      THE MOVEMENT OF OXYGEN WITHIN MUSCLES.
CC      -1- SIMILARITY: BELONGS TO THE GLOBIN FAMILY.
CC      -----
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC      or send an email to license@isb-sib.ch).

```

```

DR EMBL; V00471; CAA23743.1; -.
DR EMBL; V00472; CAA23743.1; JOINED.
DR EMBL; V00473; CAA23743.1; JOINED.
DR PIR; A02480; MYSLG.
DR PIR; A92045; MYSLH.
DR PDB; 1MSB; 30-SEP-83.
CC      DR InterPro; IPR000971; Globin.
CC      DR InterPro; IPR002335; Myoglobin.
CC      DR Pfam; PF00042; globin; 1.
CC      DR PRINTS; PR00613; MYOGLOBIN.
CC      DR PROSITE; PS01033; GLOBIN; 1.
CC      KM Heme; Oxygen transport; Transport; Muscle; 3d-structure.
CC      FT INIT_MET 0 0
CC      FT METAL 64 64 IRON (HEME DISTAL LIGAND).
CC      FT METAL 93 93 IRON (HEME PROXIMAL LIGAND).
CC      FT HELIX 3 18
CC      FT HELIX 21 35
CC      FT TURN 37 37
CC      FT TURN 38 42
CC      FT HELIX 52 56
CC      FT TURN 57 57
CC      FT HELIX 59 77
CC      FT TURN 78 78
CC      FT HELIX 86 95
CC      FT TURN 101 102
CC      FT HELIX 103 118
CC      FT TURN 124 124
CC      FT HELIX 125 149
CC      FT TURN 150 150
CC      SQ SEQUENCE 153 AA; 17297 MW; E3458330A86DE9B4 CRC64;

Query Match 5.7%; Score 6; DB 1; Length 153;
Best Local Similarity 100.0%; Pred. No. 55;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 101 GILK 106
   |||||
Db 73 GILK 78

RESULT 35
MYG_HORSE STANDARD; PRT; 153 AA.
AC P02188;
DT 21-JUL-1986 (Rel. 01, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Myoglobin.
GN MB.
OS Equus caballus (Horse), and
OS Equus burchelli (Plains zebra) (Equus quagga).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Perissodactyla; Equidae; Equus.
OC NCBI_TaxID=9796, 9790;
OX [1]
RN [1]
RP SEQUENCE.
RC SPECIES=Horse; TISSUE=Heart muscle;
RX MEDLINE=70064092; PubMed=4902609;
RA Dautrevaux M., Boulanger Y., Han K., Biserte G.;
RT "Covalent structure of horse myoglobin.";
RL Eur. J. Biochem. 11:267-277(1969).
RN [2]
RP SEQUENCE.
RC SPECIES=Horse; TISSUE=Skeletal muscle;
RA Romero-Herrera A.E., Lehmann H.;
RT "Residue 122 of sperm whale and horse myoglobin.";
RL Biochim. Biophys. Acta 336:318-323(1974).
RN [3]
RP SEQUENCE OF 1-15 AND 56-70.
RC SPECIES=Horse;
RX MEDLINE=90147691; PubMed=2302197;
RA Jahnén W., Ward L.D., Reid G.E., Moritz R.L., Simpson R.J.;
RT "Internal amino acid sequencing of proteins by in situ cyanogen

```

RT bromide cleavage in polyacrylamide gels.";  
 RL Biochem. Biophys. Res. Commun. 166:139-145(1990).  
 RN [4]  
 RP COMPOSITION OF TRYPTIC AND PEPTIC PEPTIDES.  
 RC SPECIES=E.burchelli;  
 RX MEDLINE=7518410; PubMed=1095063;  
 RA Darbe P.D., Romero-Herrera A.E., Lehmann H.;  
 RT "Comparison of the myoglobin of the zebra (*Equus burchelli*) with that  
 of the horse (*Equus caballus*).";  
 RL Biochim. Biophys. Acta 393:201-204(1975).  
 RN [5]  
 RP X-RAY CRYSTALLOGRAPHY (1.9 ANGSTROMS).  
 RC SPECIES=Horse;  
 RX MEDLINE=90294302; PubMed=2359126;  
 RA Evans S.V., Brayer G.D.;  
 RT "High-resolution study of the three-dimensional structure of horse  
 heart metmyoglobin.";  
 RL J. Mol. Biol. 213:885-897(1990).  
 RN [6]  
 RP X-RAY CRYSTALLOGRAPHY (2.8 ANGSTROMS).  
 RC SPECIES=Horse;  
 RX MEDLINE=8815746; PubMed=3346247;  
 RA Evans S.V., Brayer G.D.;  
 RT "Horse heart metmyoglobin. A 2.8-A resolution three-dimensional  
 structure determination.";  
 RL J. Biol. Chem. 263:4263-4268(1988).  
 RN [7]  
 RP X-RAY CRYSTALLOGRAPHY (1.9 ANGSTROMS) OF MUTANT THR-64.  
 RC SPECIES=Horse;  
 RX MEDLINE=95383313; PubMed=7654702;  
 RA Bogumil R., Mausus R., Hildebrand D.P., Brayer G.D., Mauk A.G.;  
 RT "Origin of the pH-dependent spectroscopic properties of  
 pentacoordinate metmyoglobin variants.";  
 RL Biochemistry 34:10483-10490(1995).  
 RN [8]  
 RP X-RAY CRYSTALLOGRAPHY (1.45 ANGSTROMS).  
 RC SPECIES=Horse;  
 RX MEDLINE=2016811; PubMed=10706294;  
 RA Chu K., Vojtechovsky J., McMahon B.H., Sweet R.M., Berendzen J.;  
 RT "Structure of a ligand-binding intermediate in wild-type carbonmonoxy  
 myoglobin.";  
 RL Nature 403:921-923(2000).  
 CC -1- FUNCTION: SERVES AS A RESERVE SUPPLY OF OXYGEN AND FACILITATES  
 THE MOVEMENT OF OXYGEN WITHIN MUSCLES.  
 CC -1- SIMILARITY: BELONGS TO THE GLOBIN FAMILY.  
 CC PIR: A02509; MYO.  
 DR PIR: A02509; MYOZ.  
 DR PIR: A34047; A34047.  
 DR PDB: 1YMA; 31-JAN-94.  
 DR PDB: 1YMB; 31-JAN-94.  
 DR PDB: 1YMC; 31-JAN-94.  
 DR PDB: 1HRM; 26-FEB-95.  
 DR PDB: 1HSY; 27-FEB-95.  
 DR PDB: 1RSE; 23-DEC-96.  
 DR PDB: 1XCH; 17-SEP-97.  
 DR PDB: 1WLA; 14-JAN-98.  
 DR PDB: 1BUE; 28-JAN-98.  
 DR PDB: 1AZI; 25-FEB-98.  
 DR PDB: 1DMR; 03-MAR-00.  
 DR PDB: 1DWS; 03-MAR-00.  
 DR PDB: 1DWT; 03-MAR-00.  
 DR InterPro: IPR000971; Globin.  
 DR InterPro: IPR002335; Myoglobin.  
 DR Pfam: PF00042; globin; 1.  
 DR PRINTS: PR00613; MYOGLOBIN.  
 DR PROSITE: PS01033; GLOBIN; 1.  
 KW Heme; Oxygen transport; Transport; Muscle; 3D-structure.  
 FT METAL 64 64 IRON (HEME DISTAL LIGAND).  
 FT METAL 93 93 IRON (HEME PROXIMAL LIGAND).  
 FT CONFLICT 122 122 D -> N (IN REF. 1).  
 FT HELIX 4 19  
 FT TURN 20 20

FT HELIX 21 35  
 FT HELIX 37 40  
 FT TURN 41 42  
 FT TURN 47 48  
 FT TURN 52 57  
 FT HELIX 59 77  
 FT TURN 78 80  
 FT HELIX 83 94  
 FT TURN 95 96  
 FT TURN 101 102  
 FT HELIX 103 118  
 FT TURN 120 122  
 FT HELIX 125 148  
 FT TURN 149 150  
 SQ SEQUENCE 153 AA; 16951 MW; 89CA01974231E93C CRC64;  
 Query Match 5.7%; Score 6; DB 1; Length 153;  
 Best Local Similarity 100.0%; Pred. No. 55;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 101 GGIILKK 106  
 Db 73 GGIILKK 78  
 RESULT 36  
 MYG\_HUMAN STANDARD; PRT; 153 AA.  
 ID MYG\_HUMAN  
 AC P02144;  
 DT 21-JUL-1986 (Rel. 01, Created)  
 DT 21-JUL-1986 (Rel. 01, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE Myoglobin.  
 GN MB.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.  
 ON NCBI\_Taxid=9606;  
 OX [1]  
 RP SEQUENCE.  
 RX MEDLINE=71291923; PubMed=5285572;  
 RA Romero-Herrera A.E., Lehmann H.;  
 RT "Primary structure of human myoglobin.";  
 RL Nature New Biol. 232:149-152(1971).  
 RN [2]  
 RP REVISIONS TO 19-22 AND 83.  
 RA Romero-Herrera A.E., Lehmann H.;  
 RT "The myoglobin of primates. I. Hylobates agilis (gibbon).";  
 RL Biochim. Biophys. Acta 251:482-488(1971).  
 RN [3]  
 RP REVISIONS TO 99-101.  
 RX MEDLINE=73002055; PubMed=4627044;  
 RA Romero-Herrera A.E., Lehmann H.;  
 RT "The myoglobin of primates. II. Pan Troglodytes (chimpanzee).";  
 RL Biochim. Biophys. Acta 278:62-67(1972).  
 RN [4]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=85232026; PubMed=2989088;  
 RA Akaboshi E.;  
 RT "Cloning of the human myoglobin gene.";  
 RL Gene 33:241-249(1985).  
 RN [5]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=84182508; PubMed=6571704;  
 RA Weller P., Jeffreys A.J., Wilson V., Blanchetot A.;  
 RT "Organization of the human myoglobin gene.";  
 RL EMBO J. 3:439-446(1984).  
 RN [6]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=20057165; PubMed=10591208;  
 RA Dunham I., Hunt A.R., Collins J.E., Bruskewich R., Beare D.M.,  
 Clamp M., Smink L.J., Atencough R., Almeida J.P., Babbage A.K.,  
 Bagguley C., Bailey J., Barlow K.F., Bates K.N., Beasley O.P.,

RA Bird C.P., Blakey S.E., Bridgeman A.M., Buck D., Burgess J.,  
 RA Burrill W.D., Burton J., Carder C., Carter N.P., Chen Y., Clark G.,  
 RA Clegg S.M., Cobley V.E., Cole C.G., Collier R.E., Connor R.,  
 RA Conroy D., Corby N.R., Coville G.J., Cox A.V., Davis J., Dawson E.,  
 RA Dhani P.D., Dockree C., Dodsorth S.J., Durbin R.M., Ellington A.G.,  
 RA Evans K.L., Fey J.M., Fleming K., French L., Garner A.A.,  
 RA Gilbert J.G.R., Goward M.E., Graham D.V., Griffiths M.N.D., Hall C.,  
 RA Hall R.E., Hall-Tamlyn G., Heathcote R.W., Ho S., Holmes S.,  
 RA Hunt S.E., Jones M.C., Kershaw J., Kimberley A.M., King A.,  
 RA Laird G.K., Langford C.F., Leverisha M.A., Lloyd C., Lloyd D.M.,  
 RA Martyn I.D., Mashreghi-Mohammadi M., Matthews L.H., Mccann O.T.,  
 RA Mcleay J., McLaren S., McMurray A.A., Milne S.A., Mortimore B.J.C.T.,  
 RA Odell C.N., Pavlt R., Pearce A.V., Pearson D., Phillimore B.J.C.T.,  
 RA Phillips S.H., Plumb R.W., Ramsey H., Ramsey Y., Rogers L.,  
 RA Scott C.E., Sehra H.K., Skuce C.D., Smalley S., Smith M.L.,  
 RA Soderlund C., Spraggon L., Steward C.A., Sulston J.E., Swann R.M.,  
 RA Vardin M., Wall M., Wallis J.M., Whiteley M.N., Wiley D.L.,  
 RA Williams L., Williams S.A., Williamson H., Wilmer T.E., Wilting L.,  
 RA Wright C.L., Hubbard T., Bentley D.R., Beck S., Rogers J., Shimizu N.,  
 RA Mitoishi S., Kawasaki K., Sasaki T., Asakawa S., Kudoh J.,  
 RA Shintani A., Shibuya K., Yoshizaki Y., Aoki N., Mitsuyama S.,  
 RA Roe B.A., Chen F., Chu L., Crabtree J., Deschamps S., Do A., Do T.,  
 RA Dorman A., Fang F., Fu Y., Hu P., Hua A., Kenton S., Lai H., Lao H.I.,  
 RA Lewis J., Lewis S., Lin S.-P., Loh P., Malaj E., Nguyen T., Pan H.,  
 RA Phan S., Qi S., Qian Y., Ray L., Ren Q., Shauli S., Sloan D., Song L.,  
 RA Wang Q., Wang Y., Wang Z., White J., Williamson D., Wu H., Yao Z.,  
 RA Zhang M., Zhang G., Chisoe S., Murray J., Miller N., Mink P.,  
 RA Fulton R., Johnson D., Bemis G., Bentley D., Bradshaw H., Bourne S.,  
 RA Cordes M., Du Z., Fulton L., Goela D., Graves T., Hawkins J.,  
 RA Hinds K., Kemp K., Latreille P., Layman D., Ozerisy P., Rohlfing T.,  
 RA Schiet P., Walker C., Wamsley A., Wohlmann P., Pepin K., Nelson J.,  
 RA Korf I., Bedell J.A., Hillier A., Mardis E., Waterston K., Wilson R.,  
 RA Emanuel B.S., Shaik T., Kurahashi H., Saitta S., Budarf M.L.,  
 RA Mcdermud H.E., Johnson A., Wong A.C.C., Morrow B.E., Edelmann L.,  
 RA Kim U.J., Shizuya H., Simon M.I., Dunanski J.P., Peyard M., Kedra D.,  
 RA Serousei E., Fransson I., Tapia I., Bruder C.E., O'Brien K.P.,  
 RA Wilkinson P., Bodenteich A., Hartman K., Hu X., Khan A.S., Lane L.,  
 RA Tiliahun Y., Wright H.,  
 RA "The DNA sequence of human chromosome 22.",  
 RL Nature 402:489-495(1999).  
 [7]  
 RN SEQUENCE OF 1-20.  
 RP TISSUE=Heart;  
 RC MEDLINE=95203287; PubMed=7895732;  
 RA Corbett J.M., Wheeler C.H., Baker C.S., Yacoub M.H., Dunn M.J.,  
 RT "The human myocardial two-dimensional gel protein database: update  
 RT 1994.",  
 RL Electrophoresis 15:1459-1465(1994).  
 [8]  
 RN VARIANT LYS-54.  
 RP MEDLINE=69275845; PubMed=5805522;  
 RA Boulton F.E., Huntsman R.G., Lorkin P.A., Lehmann H.,  
 RT "Abnormal human myoglobin: 53 (D4) glutamic acid-->lysine.",  
 RL Nature 223:832-833(1969).  
 [9]  
 RN VARIANT TRP-139.  
 RP MEDLINE=71177026; PubMed=5555219;  
 RA Boulton F.E., Huntsman R.G., Romero Herrera A., Lorkin P.A.,  
 RT Lehmann H.,  
 RA "The third variant of human myoglobin showing an unusual amino acid  
 RT substitution: 138(H16) arginine-->tryptophan.",  
 RL Biochim. Biophys. Acta 229:716-719(1971).  
 [10]  
 RN VARIANT ASN-133.  
 RP MEDLINE=71177044; PubMed=5555226;  
 RA Boulton F.E., Huntsman R.G., Romero Herrera A.E., Lorkin P.A.,  
 RT Lehmann H.,  
 RA "A human myoglobin variant 133 (H-10) lysine-->asparagine.",  
 RL Biochim. Biophys. Acta 229:871-876(1971).  
 [11]  
 RP VARIANT GLN-139.  
 MEDLINE=71085676; PubMed=5540041;  
 RX Boulton F.E., Huntsman R.G., Yawson G.I., Romero-Herrera A.E.,  
 RA Lorkin P.A.,  
 RT "The second variant of human myoglobin, 138 (H16) arginine leads to  
 RT glutamine.",  
 RL Br. J. Haematol. 20:69-74(1971).  
 [12]  
 RP X-RAY CRYSTALLOGRAPHY (2.8 ANGSTROMS) OF MUTANT WITH R-45 AND A-110.  
 RX MEDLINE=90258028; PubMed=2342104;  
 RA Hubbard S.R., Hendrickson W.A., Lambright D.G., Boxer S.G.,  
 RT "X-ray crystal structure of a recombinant human myoglobin mutant at  
 RT 2.8-A resolution.",  
 RL J. Mol. Biol. 213:215-218(1990).  
 CC -1- FUNCTION: SERVES AS A RESERVE SUPPLY OF OXYGEN AND FACILITATES  
 CC THE MOVEMENT OF OXYGEN WITHIN MUSCLES.  
 CC -1- SIMILARITY: BELONGS TO THE GLOBIN FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.1eb-sib.ch/announce/>  
 CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
 CC -----  
 DR EMBL; M14603; AAAS59595.1; -.  
 DR EMBL; M10090; AAAS59595.1; JOINED.  
 DR EMBL; M14602; AAAS59595.1; JOINED.  
 DR EMBL; X00371; CAA25109.1; -.  
 DR EMBL; X00372; CAA25109.1; JOINED.  
 DR EMBL; X00373; CAA25109.1; JOINED.  
 DR EMBL; AL049747; CAB41872.1; -.  
 DR EMBL; AL022334; CAA18457.1; -.  
 DR PIR; A02464; MYHU.  
 DR PDB; 2MM1; 15-JAN-93.  
 DR HSC-2DPAGE; P02144; HUMAN.  
 DR Genew; HGNC:6915; MB.  
 DR MIM; 160000; -.  
 DR InterPro; IPR000971; Globin.  
 DR InterPro; IPR002335; Myoglobin.  
 DR Pfam; PF00042; Globin; 1.  
 DR PRINTS; PRO0613; MYOGLOBIN.  
 DR PROSITE; PS01033; GLOBIN; 1.  
 KW Heme; Oxygen transport; Transport; Muscle; Polymorphism;  
 KW 3D-structure.  
 FT INIT MET 0 0  
 FT METAL 64 64 IRON (HEME DISTAL LIGAND).  
 FT METAL 93 93 IRON (HEME PROXIMAL LIGAND).  
 FT VARIANT 54 54 E -> K  
 FT VARIANT 133 133 /FTId=VAR\_003180.  
 FT VARIANT 139 133 K -> N.  
 FT VARIANT 139 139 /FTId=VAR\_003181.  
 FT VARIANT 139 139 R -> Q.  
 FT VARIANT 139 139 /FTId=VAR\_003182.  
 FT VARIANT 139 139 R -> W.  
 FT VARIANT 139 139 /FTId=VAR\_003183.  
 FT CONFLICT 128 128 Q -> E (IN REF. 4).  
 FT HELIX 4 17  
 FT TURN 18 19  
 FT HELIX 21 35  
 FT HELIX 37 41  
 FT TURN 42 42  
 FT TURN 45 48  
 FT HELIX 52 57  
 FT HELIX 59 76  
 FT TURN 77 80  
 FT HELIX 83 95  
 FT TURN 96 96  
 FT TURN 101 101  
 FT HELIX 102 118  
 FT HELIX 120 122  
 FT HELIX 125 148  
 FT TURN 149 150  
 SQ SEQUENCE 153 AA; 17053 MW; 5F84A2C481B8F0D5 CRC64;

Query Match 5.7%; Score 6; DB 1; Length 153;  
Best Local Similarity 100.0%; Pred. No. 55;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 101 GGLK 106  
Db 73 GGLK 78

RESULT 37  
MYG\_HYLAG STANDARD; PRT; 153 AA.  
AC P02146;  
DT 21-JUL-1986 (Rel. 01, Created)  
DT 21-JUL-1986 (Rel. 01, Last sequence update)  
DT 30-MAY-2000 (Rel. 39, Last annotation update)  
DE Myoglobin.  
GN MB.  
OS Hylobates agilis (Agile gibbon), and  
OS Hylobates syndactylus (Siamang) (Symphalangus syndactylus).  
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hylobatidae; Hylobates.  
OX NCBI\_TaxID=9579, 9590;  
RN [1]  
RP SEQUENCE.  
RC SPECIES=H.syndactylus;  
RA Romero-Herrera A.E., Lehmann H.;  
RT "The myoglobin of primates. I. Hylobates agilis (gibbon).";  
RL Biochim. Biophys. Acta 251:482-488 (1977).  
RN [2]  
RP SEQUENCE.  
RC SPECIES=H.syndactylus;  
RX MEDLINE=7720484; PubMed=672929;  
RA Bruce E.J., Casillio O., Lehmann H.;  
RT FBS Lett. 78:113-118 (1977).  
RL "The myoglobin of primates: Symphalangus syndactylus (siamang).";  
CC -1- FUNCTION: SERVES AS A RESERVE SUPPLY OF OXYGEN AND FACILITATES THE MOVEMENT OF OXYGEN WITHIN MUSCLES.  
CC -1- SIMILARITY: BELONGS TO THE GLOBIN FAMILY.  
DR PIR; A02466; MYG1.  
DR HSSP; P02144; 2MW1.  
DR InterPro; IPR000971; Globin.  
DR InterPro; IPR002335; Myoglobin.  
DR Pfam; PF00042; Globin; 1.  
DR PRINTS; PR00613; MYOGLOBIN.  
DR PROSITE; PS01033; GLOBIN; 1.  
KW Heme; Oxygen transport; Transport; Muscle.  
FT METAL 93 64 IRON (HEME DISTAL LIGAND).  
FT METAL 93 93 IRON (HEME PROXIMAL LIGAND).  
SQ SEQUENCE 153 AA; 17083 MW; 4EC0AE409C78F0D5 CRC64;

Query Match 5.7%; Score 6; DB 1; Length 153;  
Best Local Similarity 100.0%; Pred. No. 55;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 101 GGLK 106  
Db 73 GGLK 78

RESULT 38  
MYG\_INIGE STANDARD; PRT; 153 AA.  
AC P02181;  
DT 21-JUL-1986 (Rel. 01, Created)  
DT 21-JUL-1986 (Rel. 01, Last sequence update)  
DT 15-JUN-2002 (Rel. 41, Last annotation update)  
DE Myoglobin.  
GN MB.  
OS Inia geoffrensis (Amazon dolphin).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Cetartiodactyla; Cetacea; Odontoceti; Iniidae;

OC Inia.  
OX NCBI\_TaxID=9725;  
RN [1]  
RP SEQUENCE.  
RC TISSUE=Skeletal muscle;  
RX MEDLINE=76062412; PubMed=1191640.  
RA Dwalet F.E., Bogardt R.A., Jones B.N., Lehman L.D., Gurd F.R.N.;  
RT "The complete amino acid sequence of the major component myoglobin of Amazon river dolphin (Inia geoffrensis).";  
RL Biochemistry 14:5136-5143 (1975)  
CC -1- FUNCTION: SERVES AS A RESERVE SUPPLY OF OXYGEN AND FACILITATES THE MOVEMENT OF OXYGEN WITHIN MUSCLES.  
CC -1- SIMILARITY: BELONGS TO THE GLOBIN FAMILY.  
DR PIR; A02503; MYDAR.  
DR HSSP; P02188; 1MWA.  
DR InterPro; IPR000971; Globin.  
DR InterPro; IPR002335; Myoglobin.  
DR Pfam; PF00042; Globin; 1.  
DR PRINTS; PR00613; MYOGLOBIN.  
DR PROSITE; PS01033; GLOBIN; 1.  
KW Heme; Oxygen transport; Transport; Muscle.  
FT METAL 93 64 IRON (HEME DISTAL LIGAND).  
FT METAL 93 93 IRON (HEME PROXIMAL LIGAND).  
SQ SEQUENCE 153 AA; 17071 MW; 7222EDBD4BCE5D CRC64;

Query Match 5.7%; Score 6; DB 1; Length 153;  
Best Local Similarity 100.0%; Pred. No. 55;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 101 GGLK 106  
Db 73 GGLK 78

RESULT 39  
MYG\_LAGLA STANDARD; PRT; 153 AA.  
AC P02154;  
DT 21-JUL-1986 (Rel. 01, Created)  
DT 21-JUL-1986 (Rel. 01, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE Myoglobin.  
GN MB.  
OS Lagothrix lagotricha (Common woolly monkey).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Platyrrhini; Cebidae; Ateleinae;  
OC Lagothrix.  
OX NCBI\_TaxID=9519;  
RN [1]  
RP SEQUENCE.  
RC TISSUE=Skeletal muscle;  
RX MEDLINE=7321297; PubMed=4198766;  
RA Romero-Herrera A.E., Lehmann H.;  
RT "The myoglobin of primates. 4. New World monkeys: Cebidae: (1) Lagothrix lagotricha (equatorial monkey); (2) Lagothrix lagotricha (Humboldt's woolly monkey). Callitrichidae: Callitrichus jacchus (Common marmoset).";  
RL Biochim. Biophys. Acta 317:65-84 (1973).  
RN [2]  
RP PARTIAL SEQUENCE (MINOR COMPONENT).  
RA Romero-Herrera A.E., Lehmann H.;  
RT "N-terminal chain elongation as evidence for duplication of myoglobin in three South American monkeys.";  
RL FBS Lett. 31:175-180 (1973).  
CC -1- FUNCTION: SERVES AS A RESERVE SUPPLY OF OXYGEN AND FACILITATES THE MOVEMENT OF OXYGEN WITHIN MUSCLES.  
CC -1- MISCELLANEOUS: MARMOSET, WOOLLY MONKEY, AND SQUIRREL MONKEY HAVE A MINOR MYOGLOBIN COMPONENT THAT APPEARS TO DIFFER FROM EACH MAJOR COMPONENT IN HAVING PHE-1YS PRECEDING POSITION 1.  
CC -1- SIMILARITY: BELONGS TO THE GLOBIN FAMILY.  
DR PIR; A02474; MYMOW.  
DR HSSP; P02189; 1MMC.  
DR InterPro; IPR000971; Globin.

DR InterPro; IPR002335; Myoglobin.  
 DR Pfam; PF00042; Globin; 1.  
 DR PRINTS; PR00613; MYOGLOBIN.  
 DR PROSITE; PS01033; GLOBIN; 1.  
 KM Heme; Oxygen transport; Transport; Muscle.  
 FT METAL 64 64 IRON (HEME DISTAL LIGAND).  
 FT METAL 93 93 IRON (HEME PROXIMAL LIGAND).  
 FT VARIANT 1 1 G -> PKG (IN MINOR COMPONENT).  
 SQ SEQUENCE 153 AA; 17035 MW; 755C98A604ACD823 CRC64;

Query Match 5.7%; Score 6; DB 1; Length 153;  
 Best Local Similarity 100.0%; Pred. No. 55;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 101 GILKK 106  
 |||||  
 DB 73 GILKK 78

RESULT 40  
 MYG\_LAGMA STANDARD; PRT; 153 AA.  
 AC P04250; 20-MAR-1987 (Rel. 04, Created)  
 DT 20-MAR-1987 (Rel. 04, Last sequence update)  
 DT 20-MAR-1987 (Rel. 04, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Myoglobin.  
 GN MB.  
 OS Lagostomus maximus (Plains viscacha).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Hystriacognathi; Chinchillidae;  
 OC Lagostomus.  
 OC NCBI\_TaxID=10154;  
 RN [1]  
 RP SEQUENCE.  
 RA Gurnett A.M., O'Connell J.P., Harris D.E., Lehmann H., Joysey K.A.,  
 RA Nevo E.;  
 RT "The myoglobin of rodents: Lagostomus maximus (viscacha) and Spalax  
 RT ehenbergi (mole rat).";  
 RL J. Protein Chem. 3:445-454(1984).  
 CC -!- FUNCTION: SERVES AS A RESERVE SUPPLY OF OXYGEN AND FACILITATES  
 CC THE MOVEMENT OF OXYGEN WITHIN MUSCLES.  
 CC -!- SIMILARITY: BELONGS TO THE GLOBIN FAMILY.  
 CC -!- SIMILARITY: BELONGS TO THE GLOBIN FAMILY.  
 CC PIR; A02487; MYVJ.  
 DR HSSP; P02189; IMWC.  
 DR InterPro; IPR000971; Globin.  
 DR InterPro; IPR002335; Myoglobin.  
 DR Pfam; PF00042; Globin; 1.  
 DR PRINTS; PR00613; MYOGLOBIN.  
 DR PROSITE; PS01033; GLOBIN; 1.  
 KW Heme; Oxygen transport; Transport; Muscle; Polymorphism.  
 FT METAL 64 64 IRON (HEME DISTAL LIGAND).  
 FT METAL 93 93 IRON (HEME PROXIMAL LIGAND).  
 FT VARIANT 77 77 K -> R.  
 SQ SEQUENCE 153 AA; 16879 MW; C2227251D01AB448 CRC64;

Query Match 5.7%; Score 6; DB 1; Length 153;  
 Best Local Similarity 100.0%; Pred. No. 55;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 101 GILKK 106  
 |||||  
 DB 73 GILKK 78

Search completed: April 9, 2003, 13:12:58  
 Job time : 47.1967 secs



GenCore version 5.1.4 p5 4578  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 9, 2003, 13:09:27 ; Search time 45.1803 Seconds  
(without alignments)  
225.546 Million cell updates/sec

Title: US-10-020-441-2

Perfect score: 106  
Sequence: 1 VSFLLRLNGDQGGIHQPSG.....YGRDDNDPDRSKNGILKK 106

Scoring table: OLIGO  
Gapop 60.0 , Gapext 60.0

Searched: 283224 seqs, 96134422 residues

Word size : 0

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Listing first 1000 summaries

Database :

1: p1r1:\*  
2: p1r2:\*  
3: p1r3:\*  
4: p1r4:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Match	Query Length	DB ID	Description
1	106	100.0	251	2 A28942	pancreatic elastase
2	8	7.5	215	2 B69425	conserved hypotnet
3	7	6.6	150	2 B95208	conserved hypotnet
4	7	6.6	150	2 B98073	conserved hypotnet
5	7	6.6	335	2 T32657	hypothetical prote
6	7	6.6	356	2 A54032	alpha-N-acetylnaur
7	7	6.6	182	2 S41941	beta-lactamase, cl
8	7	6.6	623	2 T18250	peptide transport
9	7	6.6	656	2 B70766	hypothetical prote
10	7	6.6	686	2 B87490	NADH dehydrogenase
11	7	6.6	725	2 A45033	myelin transcripti
12	7	6.6	849	2 T46253	hypothetical prote
13	7	6.6	922	2 T06493	1,4-alpha-glucan b
14	7	6.6	1173	2 T31421	C-terminal domain
15	7	6.6	27	2 E44636	homeocic protein H
16	6	5.7	31	2 S40665	hypothetical prote
17	6	5.7	72	1 RSBPXL	excisionase - phag
18	6	5.7	72	2 S06533	excisionase - phag
19	6	5.7	72	2 A80728	excisionase [limpor
20	6	5.7	77	2 T16436	hypothetical prote
21	6	5.7	83	2 E84142	hypothetical prote
22	6	5.7	86	2 T16437	hypothetical prote
23	6	5.7	89	2 A48435	cysteine proteinas
24	6	5.7	90	2 D24356	hypothetical prote
25	6	5.7	92	2 S49385	movement protein V
26	6	5.7	95	2 B83301	hypothetical prote
27	6	5.7	95	2 H81119	conserved hypotnet
28	6	5.7	95	2 G81901	hypothetical prote
29	6	5.7	96	2 T17594	hypothetical prote

30	5.7	97	2 J00041	hypothetical prote
31	5.7	100	2 T48765	hypothetical prote
32	5.7	102	2 C84003	exogenous DNA-bind
33	5.7	106	2 T24774	hypothetical prote
34	5.7	106	2 AE2371	hypothetical prote
35	5.7	112	2 D69543	hypothetical prote
36	5.7	116	2 G89134	protein F2566.8 [i
37	5.7	122	2 G69532	conserved hypotnet
38	5.7	131	2 F97868	hypothetical prote
39	5.7	136	2 F69523	conserved hypotnet
40	5.7	145	1 R3HUI5	ribosomal protein
41	5.7	145	1 R3RT15	ribosomal protein
42	5.7	145	1 R3HY15	ribosomal protein
43	5.7	145	2 A34823	ribosomal protein
44	5.7	145	2 B34823	ribosomal protein
45	5.7	145	2 H87351	hypothetical prote
46	5.7	148	2 AH3379	arginine/ornithine
47	5.7	149	2 F82269	conserved hypotnet
48	5.7	149	2 AH0323	probable membrane
49	5.7	153	1 MYCZ	myoglobin - chimpa
50	5.7	153	1 MYGI	myoglobin - agile
51	5.7	153	1 MYGO	myoglobin - mounta
52	5.7	153	1 MYOG	myoglobin - Bornea
53	5.7	153	1 MYBAO	myoglobin - olive
54	5.7	153	1 MYMRG	myoglobin - red gu
55	5.7	153	1 MYMOH	myoglobin - hanuma
56	5.7	153	1 MYMOIM	myoglobin - crab-e
57	5.7	153	1 MYMON	myoglobin - dourou
58	5.7	153	1 MYCJ	myoglobin - common
59	5.7	153	1 MYMOC	myoglobin - brown
60	5.7	153	1 MYMQW	myoglobin - common
61	5.7	153	1 MYMOS	myoglobin - common
62	5.7	153	1 MYHH	myoglobin - wester
63	5.7	153	1 MYBD	myoglobin - Eurasi
64	5.7	153	1 MYOT	myoglobin - Eurasi
65	5.7	153	1 EN0126	myoglobin - muskra
66	5.7	153	1 MYDG	myoglobin - dog (t
67	5.7	153	1 MYFXBR	myoglobin - bat-ea
68	5.7	153	1 MYDGNH	myoglobin - Africa
69	5.7	153	1 MYFXC	myoglobin - Cape f
70	5.7	153	1 MYZC	myoglobin - Callfo
71	5.7	153	1 MYSLG	myoglobin - gray s
72	5.7	153	1 MYSLH	myoglobin [validat
73	5.7	153	1 MYBTF	myoglobin - Egypti
74	5.7	153	1 MYOY	myoglobin - aariva
75	5.7	153	1 MYTS	myoglobin - common
76	5.7	153	1 MYOL	myoglobin - Ehrenb
77	5.7	153	1 MYKS	myoglobin - casira
78	5.7	153	1 MYVJ	myoglobin - plains
79	5.7	153	1 MYLP	myoglobin - potto
80	5.7	153	1 MYLR	myoglobin - slow 1
81	5.7	153	1 MYGC	myoglobin - thick-
82	5.7	153	1 MYTG	myoglobin - weasel
83	5.7	153	1 MYOIE	myoglobin - southe
84	5.7	153	1 MYPE	myoglobin - harbor
85	5.7	153	1 MYPER	myoglobin - Dall's
86	5.7	153	1 MYDDAR	myoglobin - Amazon
87	5.7	153	1 MYWIC	myoglobin - Callfo
88	5.7	153	1 MYWHH	myoglobin - humbda
89	5.7	153	1 MYWIK	myoglobin - minke
90	5.7	153	1 MYWHR	myoglobin - finbac
91	5.7	153	1 MYWHZ	myoglobin - goose-
92	5.7	153	1 MYWHU	myoglobin - hubbs'
93	5.7	153	1 MYELI	myoglobin - Indian
94	5.7	153	1 MYELA	myoglobin - Africa
95	5.7	153	1 MYHO	myoglobin [validat
96	5.7	153	1 MYHOZ	myoglobin - common
97	5.7	153	1 MYPG	myoglobin [validat
98	5.7	153	1 MYSH	myoglobin - sheep
99	5.7	153	1 MYDE	myoglobin - red de
100	5.7	153	2 JN0344	myoglobin - Balkai
101	5.7	153	2 A61366	myoglobin - siaman
102	5.7	154	1 MYHU	myoglobin [validat









687	5	4.7	134	2	H85669	148	2	S41026	hypothetical prote
688	5	4.7	135	2	D36891	761	5	F95170	hypothetical prote
689	5	4.7	135	2	F86239	762	5	S48439	hypothetical prote
690	5	4.7	135	2	S50187	763	5	T04083	probable ribosomal
691	5	4.7	135	2	C72539	764	5	D90226	hypothetical prote
692	5	4.7	136	2	D69653	765	5	I38905	ATP-binding casset
693	5	4.7	136	2	S24090	766	5	H71661	hypothetical prote
694	5	4.7	136	2	A84775	767	5	B97807	hypothetical prote
695	5	4.7	136	2	AB0886	768	5	W6WL42	B6 protein - huma
696	5	4.7	136	2	D83828	769	5	T25581	hypothetical prote
697	5	4.7	136	2	AB1931	770	5	G64218	hypothetical prote
698	5	4.7	137	2	TA3245	771	5	H64576	ribosomal protein
699	5	4.7	137	2	S06971	772	5	AC0284	probable heat shoc
700	5	4.7	137	2	B83234	773	5	F98036	hypothetical prote
701	5	4.7	137	2	G87027	774	5	E85514	unknown protein fr
702	5	4.7	137	2	S00696	775	5	S30448	env protein - huma
703	5	4.7	137	2	H82537	776	5	S30453	env protein - huma
704	5	4.7	138	2	T49060	777	5	S30452	env protein - huma
705	5	4.7	138	2	D66715	778	5	S30450	env protein - huma
706	5	4.7	138	2	AB1449	779	5	S30451	transcription regu
707	5	4.7	139	2	A70323	780	5	A97008	ribosomal protein
708	5	4.7	139	2	E72559	781	5	E64322	50S ribosomal prot
709	5	4.7	140	1	R3HS19	782	5	S22206	unc-60 protein - C
710	5	4.7	140	1	R3HS9H	783	5	AB1879	hypothetical prote
711	5	4.7	140	1	H84321	784	5	S41727	probable exported
712	5	4.7	140	2	F72727	785	5	H75193	phosphatidylethano
713	5	4.7	140	2	A13258	786	5	AD0964	16k antigen precu
714	5	4.7	140	2	F72733	787	5	PC4216	hypothetical prote
715	5	4.7	140	2	B85059	788	5	A44947	hypothetical prote
716	5	4.7	140	2	A86313	789	5	A95023	hypothetical prote
717	5	4.7	141	1	S56403	790	5	A87264	hypothetical prote
718	5	4.7	141	1	F70457	791	5	S72685	cysteine proteinas
719	5	4.7	141	2	T15210	792	5	S57423	hypothetical prote
720	5	4.7	141	2	A91334	793	5	C64373	c-type cytochrome
721	5	4.7	141	2	B91273	794	5	G82555	ribotlavin synhas
722	5	4.7	141	2	B86114	795	5	E87359	small multidrug ex
723	5	4.7	141	2	AC0047	796	5	B75169	hypothetical prote
724	5	4.7	141	2	AG1049	797	5	D86996	hypothetical prote
725	5	4.7	141	2	AF0812	798	5	H72777	superoxide dismuta
726	5	4.7	142	2	UT0573	799	5	JQ0915	superoxide dismuta
727	5	4.7	142	2	S36509	800	5	JC1192	hypothetical prote
728	5	4.7	142	2	T18098	801	5	S23320	unknown protein, 2
729	5	4.7	142	2	T00616	802	5	A96659	conserved hypothe
730	5	4.7	142	2	JQ1776	803	5	A98062	hypothetical prote
731	5	4.7	142	2	D86796	804	5	A32200	hypothetical prote
732	5	4.7	143	2	S74747	805	5	S76220	hypothetical prote
733	5	4.7	143	2	A85072	806	5	T10014	hypothetical prote
734	5	4.7	143	2	T40617	807	5	A70700	hypothetical prote
735	5	4.7	144	2	S01391	808	5	UE0141	cyclin dependent k
736	5	4.7	144	2	B83114	809	5	B84487	hypothetical prote
737	5	4.7	144	2	D72270	810	5	B64036	hypothetical prote
738	5	4.7	144	2	S10039	811	5	G72864	AcOr-118 protein
739	5	4.7	145	1	GSYC42	812	5	C87659	conserved hypothe
740	5	4.7	145	2	C34822	813	5	H83543	hypothetical prote
741	5	4.7	145	2	F82665	814	5	T40378	hypothetical prote
742	5	4.7	146	2	A72115	815	5	T12958	copper transport p
743	5	4.7	146	2	D86508	816	5	AB1442	probable repressor
744	5	4.7	146	2	F72458	817	5	T23149	hypothetical prote
745	5	4.7	146	2	H84358	818	5	C64547	hypothetical prote
746	5	4.7	146	2	G95995	819	5	E69771	reductase, assembl
747	5	4.7	146	2	AD0929	820	5	H96925	protein containin
748	5	4.7	147	2	G86884	821	5	S05340	plastoquinol-plast
749	5	4.7	147	2	G86853	822	5	S04089	plastoquinol-plast
750	5	4.7	147	2	B90664	823	5	S16918	plastoquinol-plast
751	5	4.7	147	2	B44259	824	5	A61088	plastoquinol-plast
752	5	4.7	147	2	H85150	825	5	G70591	hypothetical prote
753	5	4.7	147	2	F98018	826	5	S22472	plastoquinol-plast
754	5	4.7	148	1	JC4649	827	5	S73261	plastoquinol-plast
755	5	4.7	148	2	AF1081	828	5	S78308	glucanokinase liamp
756	5	4.7	148	2	AG1438	829	5	AC2191	hypothetical prote
757	5	4.7	148	2	AD0483	830	5	G83300	thy-1 membrane gly
758	5	4.7	148	2	HE9813	831	5	TDRT	coat protein - sun
759	5	4.7	148	2	S41027	832	5	VCTMCP	



```
979 5 4.7 187 2 DB1130 conserved hypothet
980 5 4.7 187 2 T34628 probable acetyltra
981 5 4.7 187 2 A97605 hypothetical prote
982 5 4.7 187 2 AC2827 acetyltransferase
983 5 4.7 188 2 DB6626 aminocyl-trna hyd
984 5 4.7 188 2 DB2240 conserved hypothet
985 5 4.7 189 2 B69044 conserved hypothet
986 5 4.7 189 2 H71710 hypothetical prote
987 5 4.7 189 2 C97705 hypothetical prote
988 5 4.7 189 2 B69177 hypothetical prote
989 5 4.7 189 2 T41469 probable secretory
990 5 4.7 189 2 AG2143 hypothetical prote
991 5 4.7 190 2 JE0144 growth hormone - c
992 5 4.7 190 2 JCS682 somatotropin - com
993 5 4.7 190 2 DB4042 translation initia
994 5 4.7 190 2 T51590 membrane protein,
995 5 4.7 190 2 D70027 hypothetical prote
996 5 4.7 190 2 AH2216 nitrate transport
997 5 4.7 191 2 A97074 hypothetical prote
998 5 4.7 191 2 AE2412 hypothetical prote
999 5 4.7 191 2 AH0196 probable lipoprote
1000 5 4.7 191 2 B84207 probable acetyltra
```

## ALIGNMENTS

## RESULT 1

A28942 pancreatic elastase (EC 3.4.21.36) precursor - fluke (schistosoma mansoni)

C/Species: Schistosoma mansoni

C/Date: 30-Jun-1989 #sequence\_revision 30-Jun-1989 #text\_change 26-Aug-1999

C/Accession: A28942

R/Newport, G.R.; McKerrow, J.H.; Hedstrom, R.; Pettit, M.; McGarrigle, L.; Barr, P.J.; A  
J. Biol. Chem. 263, 13179-13184, 1988

A/Title: Cloning of the proteinase that facilitates infection by schistosome parasites.

A/Reference number: A28942; MUID:8530818; PMID:3166457

A/Accession: A28942

A/Molecule type: mRNA

A/Residues: 1-264 <NEW>

A/Cross-references: GB:J03946; NID:G160947; PID:G160948

C/Superfamily: trypsin; trypsin homology

C/Keywords: hydrolase; serine proteinase

F/27-256/Domain: trypsin homology <TRY>

Query Match 100.0%; Score 106; DB 2; Length 264;

Best Local Similarity 100.0%; Pred. No. 2e-104;

Matches 106; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VSFTLRNGDQGIHHPGSGVAVAPGYMPCMSARQRRIAGTSGFDIAIVMLAQMWL 60

DB 79 VSFTLRNGDQGIHHPGSGVAVAPGYMPCMSARQRRIAGTSGFDIAIVMLAQMWL 138

QY 61 OSGIRVVISLPSPSDIPPPCTGFIYVGYGDDNDPRPSRNGGILKK 106

DB 139 OSGIRVVISLPSPSDIPPPCTGFIYVGYGDDNDPRPSRNGGILKK 184

## RESULT 2

B69425

conserved hypothetical protein AF1403 - Archaeoglobus fulgidus

C/Species: Archaeoglobus fulgidus

C/Date: 05-Dec-1997 #sequence\_revision 05-Dec-1997 #text\_change 29-Sep-1999

C/Accession: B69425

R/Klenk, H.P.; Clayton, R.A.; Tomb, J.F.; White, O.; Nelson, K.E.; Ketchum, K.A.; Dodson  
., Fleischmann, R.D.; Quackenbush, J.; Lee, N.H.; Sutton, G.G.; Gill, S.; Kirkness, B.F.

, Glodek, A.; Zhou, L.; Overbeek, R.; Gocayne, J.D.; Weidman, J.F.; McDonald, L.

Nature 390, 364-370, 1997

A/Authors: Utecherback, T.; Cottren, M.D.; Spriggs, T.; Attiach, P.; Kaine, B.P.; Sykes, S.

Smith, H.O.; Woese, C.R.; Venter, J.C.

A/Title: The complete genome sequence of the hyperthermophilic, sulfate-reducing archae

A/Reference number: A69250; MUID:96049343; PMID:9389475

A/Accession: B69425

A/Status: preliminary; nucleic acid sequence not shown; translation not shown

A/Molecule type: DNA

A/Residues: 1-219 <KLE>

A/Cross-references: GB:AE001007; GB:AE000782; NID:G2689330; PIDN:AAB89840.1; PID:G26491

C/Superfamily: hypothetical protein MJ1458

Query Match 7.5%; Score 8; DB 2; Length 219;

Best Local Similarity 100.0%; Pred. No. 1.2;

Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 62 SGIRVISL 69

DB 168 SGIRVISL 175

## RESULT 3

B95208 conserved hypothetical protein SPI786 [imported] - Streptococcus pneumoniae (strain TIG

C/Species: Streptococcus pneumoniae

C/Date: 03-Aug-2001 #sequence\_revision 03-Aug-2001 #text\_change 03-Aug-2001

C/Accession: B95208

R/Tettelin, H.; Nelson, K.E.; Paulsen, I.T.; Eisen, J.A.; Read, T.D.; Peterson, S.; Hei  
on, J.D.; Umeyam, L.A.; White, O.; Salzberg, S.L.; Lewis, M.R.; Radune, D.; Holtapple,

nson, T.; Hickey, E.K.; Holt, I.B.

Science 293, 498-506, 2001

A/Authors: Loftus, B.J.; Yang, F.; Smith, H.O.; Venter, J.C.; Dougherty, B.A.; Morrison

A/Title: Complete Genome Sequence of a virulent isolate of Streptococcus pneumoniae.

A/Reference number: A95000; MUID:21357209; PMID:11463916

A/Accession: B95208

A/Status: preliminary

A/Molecule type: DNA

A/Residues: 1-150 <KUR>

A/Cross-references: GB:AE005672; PIDN:AAK75859.1; PID:G14973284; GSPDB:GN00164; TIGR:SP

A/Experimental source: strain TIGR4

C/Genetics:

A/Gene: SPI786

Query Match 6.6%; Score 7; DB 2; Length 150;

Best Local Similarity 100.0%; Pred. No. 10;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 69 LPSPSDI 75

DB 63 LPSPSDI 69

## RESULT 4

B98073

conserved hypothetical protein spr1612 [imported] - Streptococcus pneumoniae (strain R6

C/Species: Streptococcus pneumoniae

C/Date: 22-Oct-2001 #sequence\_revision 22-Oct-2001 #text\_change 22-Oct-2001

C/Accession: B98073

R/Hokins, J.A.; Alborn Jr., W.; Arnold, J.; Blaszcak, L.; Burgett, S.; DeHoff, B.S.;

e, R.; LeBlanc, D.J.; Lee, L.N.; Lefkowitz, E.J.; Lu, J.; Matsushima, P.; McAhren, S.;

Y, P.; Sun, P.M.; Winkler, M.E.

J. Bacteriol. 183, 5709-5717, 2001

A/Authors: Yang, Y.; Young-Bellido, M.; Zhao, G.; Zook, C.; Baltz, R.H.; Vaskunas, S.R.

A/Title: Genome of the Bacterium Streptococcus pneumoniae Strain R6.

A/Reference number: A97872; MUID:21429245; PMID:11544234

A/Accession: B98073

A/Status: preliminary

A/Molecule type: DNA

A/Residues: 1-150 <KUR>

A/Cross-references: GB:AE007317; PIDN:AAU00415.1; PID:G15459281; GSPDB:GN00174

C/Genetics:

A/Gene: spr1612

Query Match 6.6%; Score 7; DB 2; Length 150;

Best Local Similarity 100.0%; Pred. No. 10;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 69 LPSPSDI 75

DB 63 LPSPSDI 69



Db 63 LPQ9SDI 69

## RESULT 5

T32657

hypothetical protein K11D12.7 - *Caenorhabditis elegans*C/Species: *Caenorhabditis elegans*

C/Date: 29-Oct-1999 #sequence\_revision 29-Oct-1999 #text\_change 29-Oct-1999

C/Accession: T32657

R/Henkhans, J.; Wohldmann, P.; Gillam, B.

submitted to the EMBL Data Library, December 1997

A/Description: The sequence of *C. elegans* cosmid K11D12.

A/Reference number: Z21207

A/Accession: T32657

A/Status: preliminary; translated from GB/EMBL/DBJ

A/Molecule type: DNA

A/Residues: 1-335 &lt;HENS&gt;

A/Cross-references: EMBL:AF039047; PIDN:AA94227.1; GSPDB:GN00023; CESP:K11D12.7

A/Experimental source: strain Bristol N2; clone K11D12

C/Genetics:

A/Gene: CESP:K11D12.7

A/Map position: 5

A/Introns: 35/1; 88/1; 148/3; 190/3; 280/1

Query Match 6.6%; Score 7; DB 2; Length 335;  
Best Local Similarity 100.0%; Pred. No. 20;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 43 TLGGFDI 49  
|||||  
Db 111 TLGGFDI 117

## RESULT 6

A54032

alpha-N-acetylneuraminase alpha-2,8-sialyltransferase (EC 2.4.99.8) - human

N/Alternate names: G-D3 synthase; G-m3-specific alpha-2,8-sialyltransferase; ganglioside

C/Species: Homo sapiens (man)

C/Date: 02-Aug-1994 #sequence\_revision 02-Aug-1994 #text\_change 24-Sep-1999

C/Accession: A54032; I39330; I59317

R/Sasaki, K.; Kurata, K.; Kojima, N.; Kurosawa, N.; Ohta, S.; Hanai, N.; Tsuji, S.; Nish

J. Biol. Chem. 269, 15950-15956, 1994

A/Title: Expression cloning of a G-M3-specific alpha-2,8-sialyltransferase (G-D3 synthase

A/Reference number: A54032; MUID:94253194; PMID:8195250

A/Accession: A54032

A/Molecule type: mRNA

A/Residues: 1-356 &lt;SNS&gt;

A/Cross-references: GB:X77922; NID:9510987; PIDN:CAA54891.1; PID:9510988

R/Hataguchi, M.; Yamashiro, S.; Yamamoto, A.; Furukawa, K.; Takamiya, K.; Lloyd, K.O.; S

Proc. Natl. Acad. Sci. U.S.A. 91, 10455-10459, 1994

A/Title: Isolation of G03 synthase gene by expression cloning of GM3 alpha-2,8-sialyltra

A/Reference number: I39330; MUID:95024133; PMID:7937974

A/Accession: I39330

A/Molecule type: mRNA

A/Residues: 16-356 &lt;EES&gt;

A/Cross-references: GB:I32867; NID:9662253; PIDN:AAA62366.1; PID:9662254

R/Nara, K.; Matanabe, Y.; Maruyama, K.; Kasahara, K.; Nagai, Y.; Sanai, Y.

Proc. Natl. Acad. Sci. U.S.A. 91, 7952-7956, 1994

A/Title: Expression cloning of a CMP-NeuAc:NeuAc alpha2-3gal beta1-4Glc beta1-1' Cer alp

A/Reference number: I59317; MUID:9433665; PMID:8058740

A/Accession: I59317

A/Status: translated from GB/EMBL/DBJ

A/Molecule type: mRNA

A/Residues: 16-356 &lt;RE2&gt;

A/Cross-references: GB:D26360; NID:9536879; PIDN:BA05391.1; PID:9536880

C/Genetics:

A/Gene: GDB:SIAT8

A/Cross-references: GDB:370916; OMIM:601123

A/Map position: 12p12.1-12p11.2

C/Superfamily: alpha-N-acetylneuraminase alpha-2,8-sialyltransferase

C/Keywords: glycosyltransferase; transmembrane protein

Query Match 6.6%; Score 7; DB 2; Length 356;

Best Local Similarity 100.0%; Pred. No. 21;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;QY 100 NGGILKK 106  
|||||  
Db 143 NGGILKK 149

## RESULT 7

S41941

beta-lactamase, class C - *Klebsiella pneumoniae*C/Species: *Klebsiella pneumoniae*

C/Date: 06-Jan-1995 #sequence\_revision 06-Jan-1995 #text\_change 22-Jun-1999

C/Accession: S41941

R/Gonzalez-Leiza, M.; Perez-Diaz, J.C.; Casellas, J.M.; Martinez-Beltran, J.; Bush, K.

submitted to the EMBL Data Library, January 1994

A/Description: Gene sequence and biochemical characterization of FOX-1 from *Klebsiella*

A/Reference number: S41941

A/Accession: S41941

A/Status: preliminary

A/Molecule type: DNA

A/Residues: 1-382 &lt;GON&gt;

A/Cross-references: EMBL:X77455; NID:9453214; PIDN:CAA54602.1; PID:9453215

C/Superfamily: Escherichia coli beta-lactamase

Query Match 6.6%; Score 7; DB 2; Length 382;  
Best Local Similarity 100.0%; Pred. No. 22;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 49 IAIWLA 55  
|||||  
Db 354 IAIWLA 360

## RESULT 8

T18250

peptide transport protein - yeast (*Candida albicans*)C/Species: *Candida albicans*

C/Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 15-Oct-1999

C/Accession: T18250

R/Barrell, B.G.; Rajandream, M.A.

submitted to the EMBL Data Library, November 1998

A/Reference number: Z18831

A/Accession: T18250

A/Status: preliminary; translated from GB/EMBL/DBJ

A/Molecule type: DNA

A/Residues: 1-623 &lt;BAR&gt;

A/Cross-references: EMBL:AL033503; NID:e1341066; PID:e1341077; PIDN:CAA20201.1

C/Genetics:

A/Gene: PTR2

A/Introns: 475/3

Query Match 6.6%; Score 7; DB 2; Length 623;  
Best Local Similarity 100.0%; Pred. No. 33;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 50 AIWMLAQ 56  
|||||  
Db 572 AIWMLAQ 578

## RESULT 9

B70766

hypothetical protein Rv2079 - *Mycobacterium tuberculosis* (strain H37RV)C/Species: *Mycobacterium tuberculosis*

C/Date: 17-Jul-1998 #sequence\_revision 17-Jul-1998 #text\_change 22-Oct-1999

C/Accession: B70766

R/Cole, S.T.; Broesch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon,

Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.

Nature 393, 537-544, 1998

A/Authors: Squares, R.; Sultcon, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.

A/Title: Deciphering the biology of *Mycobacterium tuberculosis* from the complete genome

A:Reference number: A70500; MUID:98295987; PMID:9634230  
A:Accession: B70766  
A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-656 <COL>  
A:Cross-references: GB:Z73966; GB:AL123456; NID:g3261577; PIDN:CAA98192.1; PID:e247004;  
A:Experimental source: strain H37Rv  
C:Genetics:  
A:Gene: RV2079

Query Match 6.6%; Score 7; DB 2; Length 656;  
Best Local Similarity 100.0%; Pred. No. 35;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 75 IPPPGTG 81  
|||||  
Db 222 IPPPGTG 228

RESULT 10  
E87490  
NADH dehydrogenase I, G subunit CC1946 [imported] - Caulobacter crescentus  
C:Species: Caulobacter crescentus  
C:Date: 20-Apr-2001 #sequence\_revision 20-Apr-2001 #text\_change 10-May-2001  
C:Accession: E87490  
R:Merlan, W.C.; Feldblyum, T.V.; Paulsen, I.T.; Nelson, K.E.; Eisen, J.; Heidelberg, J.  
B.; Laub, M.T.; DeBoy, R.T.; Dodson, R.J.; Durkin, A.S.; Gwinn, M.L.; Haft, D.H.; Kolon  
n, J.; Ermolaeva, M.; White, O.; Salzberg, S.L.; Shapiro, L.; Venter, J.C.; Fraser, C.M.  
Proc. Natl. Acad. Sci. U.S.A. 98, 4136-4141, 2001  
A:Title: Complete Genome Sequence of Caulobacter crescentus.  
A:Reference number: A87249; MUID:21173698; PMID:11259647  
A:Accession: E87490  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-686 <STO>  
A:Cross-references: GB:AEO05673; NID:g13423405; PIDN:AAK23921.1; GSPDB:GN00148  
C:Genetics:  
A:Gene: CC1946  
C:Superfamily: NADH dehydrogenase (ubiquinone) chain 11

Query Match 6.6%; Score 7; DB 2; Length 686;  
Best Local Similarity 100.0%; Pred. No. 36;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 85 VGYGRD 91  
|||||  
Db 115 VGYGRD 121

RESULT 11  
A45033  
myelin transcription factor 1 - human (fragment)  
N:Alternate names: proteolipid protein binding protein; transcription factor MTF1  
C:Species: Homo sapiens (man)  
C:Date: 10-Jun-1993 #sequence\_revision 18-Nov-1994 #text\_change 05-Nov-1999  
C:Accession: A45033; S27964  
R:Kim, J.G.; Hudson, L.D.  
Mol. Cell. Biol. 12, 5632-5639, 1992  
A:Title: Novel member of the zinc finger superfamily: A C2-HC finger that recognizes a g  
A:Reference number: A45033; MUID:93078764; PMID:1280325  
A:Accession: A45033  
A:Status: preliminary; not compared with conceptual translation  
A:Molecule type: mRNA  
A:Residues: 1-725 <KIM>  
A:Cross-references: EMBL:M96980; NID:g189041; PIDN:AAA59897.1; PID:g189042  
A:Experimental source: fetal brain  
A>Note: sequence extracted from NCBI Backbone (NCBI:119201)  
C:Genetics:  
A:Gene: GDB:MTF1; PLPBI  
A:Cross-references: GDB:127811  
C:Keywords: DNA binding; transcription factor; zinc finger  
Query Match 6.6%; Score 7; DB 2; Length 725;

Best Local Similarity 100.0%; Pred. No. 38;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 19 SGVKVAP 25  
|||||  
Db 478 SGVKVAP 484

RESULT 12  
T46253  
hypothetical protein DKFZp761f171.1 - human (fragment)  
C:Species: Homo sapiens (man)  
C:Date: 04-Feb-2000 #sequence\_revision 04-Feb-2000 #text\_change 04-Feb-2000  
C:Accession: T46253  
R:Ottewill, B.; Obermaier, B.; Mewes, H.W.; Gaassenhuber, J.; Wiemann, S.  
submitted to the Protein Sequence Database, January 2000  
A:Reference number: Z23031  
A:Accession: T46253  
A:Status: preliminary  
A:Molecule type: mRNA  
A:Residues: 1-849 <AAA>  
A:Cross-references: EMBL:AL137500  
A:Experimental source: adult amygdala; clone DKFZp761f171  
C:Genetics:  
A>Note: DKFZp761f171.1

Query Match 6.6%; Score 7; DB 2; Length 849;  
Best Local Similarity 100.0%; Pred. No. 43;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 57 MVNIQSG 63  
|||||  
Db 126 MVNIQSG 132

RESULT 13  
T06493  
1,4-alpha-glucan branching enzyme (EC 2.4.1.18) I - garden pea  
N:Alternate names: starch branching enzyme I  
C:Species: Pisum sativum (garden pea)  
C:Date: 23-Apr-1999 #sequence\_revision 23-Apr-1999 #text\_change 21-Jul-2000  
C:Accession: T06493  
R:Buton, R.A.; Bewley, J.D.; Smith, A.M.; Bhattacharyya, M.K.; Tatge, H.; Ring, S.; Bu  
plant J. 7, 3-15, 1995  
A:Title: Starch branching enzymes belonging to distinct enzyme families are differentia  
A:Reference number: Z15717; MUID:95201826; PMID:7894509  
A:Accession: T06493  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: mRNA  
A:Residues: 1-922 <BUR>  
A:Cross-references: EMBL:X80009; NID:g510545; PIDN:CAA56319.1; PID:g1345570  
C:Genetics:  
A:Gene: SBEI  
C:Superfamily: 1,4-alpha-glucan branching enzyme  
C:Keywords: chloroplast; glycogen/starch biosynthesis; glycoeyltransferase; hexosyltran

Query Match 6.6%; Score 7; DB 2; Length 922;  
Best Local Similarity 100.0%; Pred. No. 46;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 75 IPPPGTG 81  
|||||  
Db 159 IPPPGTG 165

RESULT 14  
T31421  
C-terminal domain-binding protein rAI - rat  
C:Species: Rattus norvegicus (Norway rat)  
C:Date: 29-Oct-1999 #sequence\_revision 29-Oct-1999 #text\_change 07-Dec-1999  
C:Accession: T31421  
R:Yuryev, A.; Patutajan, M.; Litington, Y.; Joshi, R.V.; Gentile, C.; Gebara, M.; Cor  
Proc. Natl. Acad. Sci. U.S.A. 93, 6975-6980, 1996

A>Title: The C-terminal domain of the largest subunit of RNA polymerase II interacts with  
A/Reference number: 221024; MUID:96293459; PMID:8629292  
A/Accession: T31421  
A/Status: preliminary; translated from GB/EMBL/DBJ  
A/Molecule type: mRNA  
A/Residues: 1-1173 <YR>  
A/Cross-references: EMBL:U49056; NID:g1438531; PID:g1438532; PIDN:AMC52657.1  
A/Experimental source: hippocampus

Query Match 6.6%; Score 7; DB 2; Length 1173;  
Best Local Similarity 100.0%; Pred. No. 57;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 20 GVKVAPG 26  
|||||  
Db 751 GVKVAPG 757

RESULT 15  
E44636  
homotetic protein Hox 4 (clone 4d) - acorn worm (fragment)  
C/Species: Saccoglossus kowalevskii  
C/Date: 19-Mar-1997 #sequence\_revision 19-Dec-1997 #text\_change 15-Oct-1999  
C/Accession: E44636  
R/pendleton, J.W.; Nagai, B.K.; Murtha, M.T.; Ruddle, F.H.  
Proc. Natl. Acad. Sci. U.S.A. 90, 6300-6304, 1993  
A/Title: Expansion of the Hox gene family and the evolution of chordates.  
A/Reference number: A44616; MUID:93317669; PMID:8101001

A/Accession: E44636  
A/Status: preliminary  
A/Molecule type: DNA  
A/Residues: 1-27 <PEN>  
A/Cross-references: GB:U4915; NID:9294800; PIDN:AAA02550.1; PID:9294801  
C/Superfamily: unassigned homeobox proteins; homeobox homology  
C/Keywords: DNA binding; embryo; homeobox; nucleus; transcription regulation

Query Match 5.7%; Score 6; DB 2; Length 27;  
Best Local Similarity 100.0%; Pred. No. 28;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 55 AQMVNL 60  
|||||  
Db 15 AQMVNL 20

RESULT 16  
S40665  
hypothetical protein - wheat dwarf virus  
C/Species: wheat dwarf virus  
C/Date: 19-Mar-1997 #sequence\_revision 19-Mar-1997 #text\_change 17-Mar-1999  
C/Accession: S40665  
R/Dekker, E.L.; Woolston, C.J.; Xue, Y.; Cox, B.; Mullineaux, P.M.  
Nucleic Acids Res. 19, 4075-4081, 1991  
A/Title: Transcription mapping reveals different expression strategies for the bicistronic  
A/Reference number: S40665; MUID:91334111; PMID:1870964  
A/Accession: S40665  
A/Status: preliminary  
A/Molecule type: genomic RNA  
A/Residues: 1-31 <DEK>

Query Match 5.7%; Score 6; DB 2; Length 31;  
Best Local Similarity 100.0%; Pred. No. 31;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 8 NGDOOG 13  
|||||  
Db 6 NGDOOG 11

RESULT 17  
RSBPXL  
excisionase - phage lambda  
C/Species: phage lambda

C/Date: 31-Oct-1980 #sequence\_revision 23-Oct-1981 #text\_change 23-Jul-1999  
C/Accession: C94164; C43012; A93699; A93844; A04321  
R/Daniels, D.  
submitted to the Nucleic Acid Sequence Database, September 1982  
A/Reference number: A94614  
A/Accession: C94164

A/Molecule type: DNA  
A/Residues: 1-72 <DAN>  
R/Sanger, F.; Coulson, A.R.; Hong, G.F.; Hill, D.F.; Petersen, G.B.  
J. Mol. Biol. 162, 729-773, 1982  
A/Title: Nucleotide sequence of bacteriophage lambda DNA.  
A/Reference number: A92891; MUID:85189071; PMID:6221115  
A/Accession: C43012

A/Molecule type: DNA  
A/Residues: 1-72 <SAN>  
A/Cross-references: GB:U02459; GB:M17233; GB:M24325; GB:V00636; GB:X00906; NID:g215104;  
R/Davies, R.W.  
Nucleic Acids Res. 8, 1765-1782, 1980  
A/Title: DNA sequence of the int-xis P-I region of the bacteriophage lambda; overlap o

A/Reference number: A93699; MUID:81053845; PMID:6253947  
A/Accession: A93699  
A/Molecule type: DNA

A/Residues: 1-72 <DAV>  
A/Cross-references: GB:U02459; GB:M17233; GB:M24325; GB:V00636; GB:X00906; NID:g215104;  
R/Hoess, R.H.; Foeller, C.; Bidwell, K.; Landy, A.  
Proc. Natl. Acad. Sci. U.S.A. 77, 2482-2486, 1980  
A/Title: Site-specific recombination functions of bacteriophage lambda: DNA sequence of  
A/Reference number: A93844; MUID:80234646; PMID:6446773

A/Accession: A93844  
A/Molecule type: DNA  
A/Residues: 1-72 <HOE>  
A/Cross-references: GB:U02459; GB:M17233; GB:M24325; GB:V00636; GB:X00906; NID:g215104;  
C/Comment: Excisionase and lambda integrase are necessary for the excision of prophage  
C/Genetics:

A/Gene: xis  
A/Map position: 59.95-59.51  
C/Superfamily: phage lambda excisionase  
C/Keywords: DNA binding

Query Match 5.7%; Score 6; DB 1; Length 72;  
Best Local Similarity 100.0%; Pred. No. 63;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 34 ARQRRP 39  
|||||  
Db 10 ARQRRP 15

RESULT 18  
S06533  
excisionase - phage HK022  
C/Species: phage HK022  
C/Date: 03-May-1994 #sequence\_revision 20-Feb-1995 #text\_change 20-Sep-1999  
C/Accession: S06533  
R/Yeglig, E.; Dolev, S.; Oberto, J.; Kislav, N.; Ramalah, N.; Weisberg, R.A.  
J. Mol. Biol. 207, 695-717, 1989  
A/Title: Determinants of site-specific recombination in the lambdaoid coliphage HK022. A

A/Reference number: S04590; MUID:9342457; PMID:2547971  
A/Accession: S06533  
A/Status: translation not shown  
A/Molecule type: DNA  
A/Residues: 1-72 <YAG>  
A/Cross-references: EMBL:X51962; NID:g15759; PIDN:CMA36222.1; PID:g15761  
C/Genetics:  
A/Gene: xis  
C/Superfamily: phage lambda excisionase  
C/Keywords: DNA binding

Query Match 5.7%; Score 6; DB 2; Length 72;  
Best Local Similarity 100.0%; Pred. No. 63;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 34 ARQRRP 39

|||||  
Db 10 ARORRP 15

## RESULT 19

A90729  
excisionase [imported] - Escherichia coli (strain O157:H7, substrain R1MD 0509952)  
C/Species: Escherichia coli  
C/Date: 18-Jul-2001 #sequence\_revision 18-Jul-2001 #text\_change 17-May-2002  
C/Accession: A90729  
R/Hayashi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.G.  
gasawara, N.; Yasunaga, T.; Kuwara, S.; Shiba, T.; Hattori, M.; Shinagawa, H.  
DNA Res. 8, 11-22, 2001  
A/Title: Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and gene  
A/Reference number: A96629; MUID:21156231; PMID:11258796  
C/Accession: A90729  
A/Status: preliminary  
A/Molecule type: DNA  
A/Residues: 1-72 <HAY>  
A/Cross-references: GB:BA000007, PIDN:BA834224.1; PID:G13360260; GSPDB:GN00154  
A/Experimental source: strain O157:H7, substrain R1MD 0509952  
C/Genetics:  
A/Gene: EC60801  
C/Superfamily: phage lambda excisionase

Query Match 5.7%; Score 6; DB 2; Length 72;  
Best Local Similarity 100.0%; Pred. No. 63;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 34 ARORRP 39  
|||||  
Db 10 ARORRP 15

## RESULT 20

T16436  
hypotheical protein F53A9.1 - Caenorhabditis elegans  
C/Species: Caenorhabditis elegans  
C/Date: 20-Sep-1999 #sequence\_revision 20-Sep-1999 #text\_change 20-Sep-1999  
C/Accession: T16436  
R/Miller, N.  
submitted to the EMBL Data Library, March 1995  
A/Description: The sequence of C. elegans cosmid F53A9.  
A/Reference number: Z18513  
A/Accession: T16436  
A/Status: preliminary; translated from GB/EMBL/DBJ  
A/Molecule type: DNA  
A/Residues: 1-77 <MIL>  
A/Cross-references: EMBL:U23523; NID:G746551; PID:G746552; PIDN:AA046556.1; CESP:F53A9.1  
A/Experimental source: strain Bristol N2  
C/Genetics:  
A/Gene: CESP:F53A9.1

Query Match 5.7%; Score 6; DB 2; Length 77;  
Best Local Similarity 100.0%; Pred. No. 67;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 24 APGYMP 29  
|||||  
Db 25 APGYMP 30

## RESULT 21

E84142  
hypotheical protein BH3941 [imported] - Bacillus halodurans (strain C-125)  
C/Species: Bacillus halodurans  
C/Date: 01-Dec-2000 #sequence\_revision 01-Dec-2000 #text\_change 15-Jun-2001  
C/Accession: E84142  
R/Takami, H.; Nakasone, K.; Takaki, Y.; Maeno, G.; Sasaki, R.; Masui, N.; Fujii, F.; Hire  
Nucleic Acids Res. 28, 4317-4331, 2000  
A/Title: Complete genome sequence of the alkaliphilic bacterium Bacillus halodurans and  
A/Reference number: A83650; MUID:20512582; PMID:11058132  
A/Accession: E84142

A/Status: preliminary  
A/Molecule type: DNA  
A/Residues: 1-83 <STO>  
A/Cross-references: GB:AP001520; GB:BA000004; NID:G10176401; PIDN:BA807660.1; GSPDB:GN0  
A/Experimental source: strain C-125  
C/Genetics:  
A/Gene: BH3941

Query Match 5.7%; Score 6; DB 2; Length 83;  
Best Local Similarity 100.0%; Pred. No. 71;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 21 VKVAPG 26  
|||||  
Db 56 VKVAPG 61

## RESULT 22

T16437  
hypotheical protein F53A9.6 - Caenorhabditis elegans  
C/Species: Caenorhabditis elegans  
C/Date: 20-Sep-1999 #sequence\_revision 20-Sep-1999 #text\_change 20-Sep-1999  
C/Accession: T16437  
R/Miller, N.  
submitted to the EMBL Data Library, March 1995  
A/Description: The sequence of C. elegans cosmid F53A9.  
A/Reference number: Z18513  
A/Accession: T16437  
A/Status: preliminary; translated from GB/EMBL/DBJ  
A/Molecule type: DNA  
A/Residues: 1-86 <MIL>  
A/Cross-references: EMBL:U23523; NID:G746551; PID:G746557; PIDN:AA046561.1; CESP:F53A9.  
A/Experimental source: strain Bristol N2  
C/Genetics:  
A/Gene: CESP:F53A9.6

Query Match 5.7%; Score 6; DB 2; Length 86;  
Best Local Similarity 100.0%; Pred. No. 73;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 24 APGYMP 29  
|||||  
Db 27 APGYMP 32

## RESULT 23

A48435  
cysteine proteinase AC-3 - nematode (Haemonchus contortus) (fragment)  
C/Species: Haemonchus contortus  
C/Date: 01-Dec-1993 #sequence\_revision 18-Nov-1994 #text\_change 22-Jun-1999  
C/Accession: A48435  
R/Parr, D.; Ames, L.G.; Hageman, R.; Reynolds, V.; Boisvenue, R.J.; Cox, G.N.  
Mol. Biochem. Parasitol. 51, 209-218, 1992  
A/Title: Cloning and sequence comparisons of four distinct cysteine proteases expressed  
A/Reference number: A48435; MUID:92244291; PMID:15744079  
A/Accession: A48435  
A/Status: preliminary  
A/Molecule type: DNA  
A/Residues: 1-89 <PRA>  
A/Cross-references: GB:M80393; NID:G159180; PIDN:AAA29179.1; PID:G552147  
A/Note: sequence extracted from NCBI backbone (NCBIN:98508, NCBI:P:98519)  
C/Superfamily: papain

Query Match 5.7%; Score 6; DB 2; Length 89;  
Best Local Similarity 100.0%; Pred. No. 75;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 41 AQTLSG 46  
|||||  
Db 30 AQTLSG 35

## RESULT 24

D24356  
hypothetical protein, 10K - wheat dwarf virus  
C/Species: wheat dwarf virus  
C/Date: 10-Mar-1994 #sequence\_revision 10-Mar-1994 #text\_change 08-Oct-1999  
C/Accession: D24356  
R/McDonald, S.W.; MacDonald, H.; Hamilton, W.D.O.; Coutts, R.H.A.; Buck, K.W.  
EMBO J. 4, 2173-2180, 1985  
A/Title: The nucleotide sequence of cloned wheat dwarf virus DNA.  
A/Reference number: A91012  
A/Accession: D24356  
A/Status: translation not shown  
A/Molecule type: DNA  
A/Residues: 1-90 <MAC>  
A/Cross-references: GB:X02869; NID:G59382; PIDN:CAA26621.1; PID:G59383

Query Match  
Best Local Similarity 5.7%; Score 6; DB 2; Length 90;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 8 NGDQOG 13  
|||||  
DB 82 NGDQOG 87

RESULT 25  
S49385  
movement protein VI - wheat dwarf virus  
C/Species: wheat dwarf virus  
C/Date: 16-Feb-1995 #sequence\_revision 12-May-1995 #text\_change 08-Oct-1999  
C/Accession: S49385  
R/Bendaiane, M.; Schalk, H.J.; Gronenborn, B.  
Submitted to the EMBL Data Library, October 1994  
A/Description: Identification and characterization of wheat dwarf virus (WDV) from France  
A/Reference number: S49385  
A/Accession: S49385  
A/Status: preliminary  
A/Molecule type: DNA  
A/Residues: 1-92 <BEN>  
A/Cross-references: EMBL:X82104; NID:G557674; PIDN:CAA57623.1; PID:G557675

Query Match  
Best Local Similarity 5.7%; Score 6; DB 2; Length 92;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 8 NGDQOG 13  
|||||  
DB 84 NGDQOG 89

RESULT 26  
B83301  
hypothetical protein PA2747 [imported] - Pseudomonas aeruginosa (strain PA01)  
C/Species: Pseudomonas aeruginosa  
C/Date: 15-Sep-2000 #sequence\_revision 15-Sep-2000 #text\_change 31-Dec-2000  
C/Accession: B83301  
R/Stover, C.K.; Plam, X.Q.; Eryin, A.L.; Mizoguchi, S.D.; Warren, P.; Hickey, M.J.; Boman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Lathig, K.; Lim, J.; Lory, S.; Olson, M.V.  
Nature 406, 959-964, 2000  
A/Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic pathogen  
A/Reference number: A82950; MUID:20437337; PMID:10984043  
A/Accession: B83301  
A/Status: preliminary  
A/Molecule type: DNA  
A/Residues: 1-95 <STO>  
A/Cross-references: GB:AE004703; GB:AE004091; NID:G9948825; PIDN:AG06135.1; GSPDB:GN001  
A/Experimental source: strain PA01  
C/Genetics:  
A/Gene: PA2747

Query Match  
Best Local Similarity 5.7%; Score 6; DB 2; Length 95;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4 LTRNG 9  
|||||  
DB 30 LTRNG 35

RESULT 27  
H8119  
conserved hypothetical protein NMB118 [imported] - Neisseria meningitidis (strain MC58)  
C/Species: Neisseria meningitidis  
C/Date: 31-Mar-2000 #sequence\_revision 31-Mar-2000 #text\_change 19-Jan-2001  
C/Accession: H8119  
R/Tectelin, H.; Saunders, N.J.; Heidelberg, J.; Jeffries, A.C.; Nelson, K.E.; Eisen, J. Hickey, E.K.; Haft, D.H.; Salzberg, S.L.; White, O.; Fleischmann, R.D.; Dougherty, B.A. et al.; H. Qiu, H.; Vamathevan, J.; Gill, J.; Scarlato, V.; Masiatani, V.; Piazza, M.  
Science 287, 1809-1815, 2000  
A/Authors: Grandi, G.; Sun, L.; Smith, H.O.; Fraser, C.M.; Moxon, E.R.; Rappuoli, R.; V.  
A/Title: Complete genome sequence of Neisseria meningitidis serogroup B strain MC58.  
A/Reference number: A81000; MUID:20175755; PMID:10710307  
A/Accession: H8119  
A/Status: preliminary  
A/Molecule type: DNA  
A/Residues: 1-95 <TET>  
A/Cross-references: GB:AE002461; GB:AE002098; NID:G7226350; PIDN:AAF41508.1; PID:G72263  
A/Experimental source: serogroup B, strain MC58  
C/Genetics:  
A/Gene: NMB118

Query Match  
Best Local Similarity 5.7%; Score 6; DB 2; Length 95;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 63 GIRVIS 68  
|||||  
DB 70 GIRVIS 75

RESULT 28  
G81901  
hypothetical protein NMA1327 [imported] - Neisseria meningitidis (strain Z2491 serogroup G)  
C/Species: Neisseria meningitidis  
C/Date: 05-May-2000 #sequence\_revision 05-May-2000 #text\_change 02-Feb-2001  
C/Accession: G81901  
R/Parkhill, J.; Achtman, M.; James, K.D.; Bentley, S.D.; Churcher, C.; Klee, S.R.; More; Holroyd, S.; Jorgensen, K.; Leather, S.; Moule, S.; Mungall, K.; Quail, M.A.; Rajandream  
Nature 404, 502-506, 2000  
A/Title: Complete DNA sequence of a serogroup A strain of Neisseria meningitidis Z2491.  
A/Reference number: A81775; MUID:2022556; PMID:10761919  
A/Accession: G81901  
A/Status: preliminary  
A/Molecule type: DNA  
A/Residues: 1-95 <PAR>  
A/Cross-references: GB:AL162755; GB:AL157959; NID:G7379742; PIDN:CAM84578.1; PID:G73800  
A/Experimental source: serogroup A, strain Z2491  
C/Genetics:  
A/Gene: NMA1327

Query Match  
Best Local Similarity 5.7%; Score 6; DB 2; Length 95;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 63 GIRVIS 68  
|||||  
DB 70 GIRVIS 75

RESULT 29  
T17594  
hypothetical protein a104L - Chlorella virus PBCV-1  
C/Species: Chlorella virus PBCV-1  
C/Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 15-Oct-1999  
C/Accession: T17594  
R/Graves, M.V.; Van Etten, J.L.

submitted to the EMBL Data Library, May 1999  
A:Reference number: Z18806  
A:Accession: T17594  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-96 <GRA>  
A:Cross-references: EMBL:U42580, NID:G4028896, PIDN:AAC96472.1  
A:Experimental source: specific host *Chlorella* strain NC64A  
C:Genetics:  
A>Note: al04L

Query Match 5.7%; Score 6; DB 2; Length 96;  
Best Local Similarity 100.0%; Pred. No. 80;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 99 KNKGIL 104  
Db 57 KNKGIL 62

RESULT 30  
JQ0041  
hypothetical protein, 11K - *Escherichia coli* insertion sequence IS2  
C:Species: *Escherichia coli*  
C:Date: 31-Mar-1990 #sequence\_revision 31-Mar-1990 #text\_change 08-Oct-1999  
C:Accession: JQ0041  
R:Ronecker, H.J.; Rak, B.  
Gene 59, 291-296, 1987  
A:Title: Genetic organization of insertion element IS2 based on a revised nucleotide seq  
A:Reference number: A91582; MUID:88137965; PMID:2830172  
A:Accession: JQ0041  
A:Status: translation not shown  
A:Molecule type: DNA  
A:Residues: 1-97 <RON>  
A:Cross-references: GB:V00279; GB:J01732; GB:M19426; NID:G41525; PIDN:CAA23540.1; PID:95  
A:Experimental source: strain K-12, strain HB101  
C:Genetics:  
A:Mobile element: insertion sequence IS2  
A:Start codon: GTG

Query Match 5.7%; Score 6; DB 2; Length 97;  
Best Local Similarity 100.0%; Pred. No. 81;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 84 IVGYGR 89  
Db 44 IVGYGR 49

RESULT 31  
T48765  
hypothetical protein 13E11.130 [imported] - *Neurospora crassa*  
C:Species: *Neurospora crassa*  
C:Date: 05-May-2000 #sequence\_revision 05-May-2000 #text\_change 19-May-2000  
C:Accession: T48765  
R:Schule, U.; Aizm, V.; Hohnesiel, J.; Brandt, P.; Farfmann, B.; Holland, R.; Nyakatura,  
submitted to the Protein Sequence Database, April 2000  
A:Reference number: Z24541  
A:Accession: T48765  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-100 <SCH>  
A:Cross-references: EMBL:AL353820; GSPDB:GN00112; NCSP:13E11.130  
A:Experimental source: cosmid contig 13E11; strain 74  
C:Genetics:  
A:Gene: NCSP:13E11.130  
A:Map position: 2  
C:Superfamily: *Neurospora crassa* hypothetical protein 13E11.130

Query Match 5.7%; Score 6; DB 2; Length 100;  
Best Local Similarity 100.0%; Pred. No. 83;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 43 TLGGFD 48  
Db 47 TLGGFD 52

RESULT 32  
C84003  
exogenous DNA-binding protein comGC [imported] - *Bacillus halodurans* (strain C-125)  
C:Species: *Bacillus halodurans*  
C:Date: 01-Dec-2000 #sequence\_revision 01-Dec-2000 #text\_change 15-Jun-2001  
C:Accession: C84003  
R:Ikami, H.; Nakasone, K.; Takaki, Y.; Maeno, G.; Sasaki, R.; Masui, N.; Fuji, F.; Hir  
Nucleic Acids Res. 28, 4317-4331, 2000  
A:Title: Complete genome sequence of the alkaliphilic bacterium *Bacillus halodurans* and  
A:Reference number: A83650; MUID:20512582; PMID:11058132  
A:Accession: C84003  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-102 <STO>  
A:Cross-references: GB:AP001516; GB:BA000004; NID:G10175192; PIDN:BAB06546.1; GSPDB:GN0  
A:Experimental source: strain C-125  
C:Genetics:  
A:Gene: comGC

Query Match 5.7%; Score 6; DB 2; Length 102;  
Best Local Similarity 100.0%; Pred. No. 85;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 4 LTRNG 9  
Db 91 LTRNG 96

RESULT 33  
T24774  
hypothetical protein T10B10.9 - *Caenorhabditis elegans*  
C:Species: *Caenorhabditis elegans*  
C:Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 18-Feb-2000  
C:Accession: T24774  
R:Sims, M.  
submitted to the EMBL Data Library, May 1996  
A:Reference number: Z19934  
A:Accession: T24774  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-106 <WIL>  
A:Cross-references: EMBL:Z72514; PIDN:CAB54309.1; GSPDB:GN00028; CESP:T10B10.9  
A:Experimental source: clone T10B10  
C:Genetics:  
A:Gene: CESP:T10B10.9  
A:Map position: X  
A:introns: 57/1

Query Match 5.7%; Score 6; DB 2; Length 106;  
Best Local Similarity 100.0%; Pred. No. 87;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 100 NGGIIL 105  
Db 11 NGGIIL 16

RESULT 34  
AE2371  
hypothetical protein alr4525 [imported] - *Nostoc* sp. (strain PCC 7120)  
C:Species: *Nostoc* sp.  
A>Note: *Nostoc* sp. strain PCC 7120 is a synonym of *Anabaena* sp. strain PCC 7120  
C:Date: 14-Dec-2001 #sequence\_revision 14-Dec-2001 #text\_change 30-Jun-2002  
C:Accession: AE2371  
R:Kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.; Iriuch  
Nakazaki, N.; Shimura, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Yasuda, M.; Tabata,  
DNA Res. 8, 205-213, 2001  
A:Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing *Cyanobacterium* An

A:Reference number: AB1807; MUID:21595285; PMID:11759840  
A:Accession: AE2371  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-106 <KUR>  
A:Cross-references: GB:BA000019; PIDN:BA876224.1; PID:G1713361; GSPDB:GN00179  
A:Experimental source: strain FCC 7120  
C:Genetics:  
A:Gene: alx4525

Query Match 5.7%; Score 6; DB 2; Length 106;  
Best Local Similarity 100.0%; Pred. No. 87;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 81 GVFIVG 86  
|||  
Db 20 GVFIIVG 25

RESULT 35  
D69543  
hypothetical protein AF2348 - Archaeoglobus fulgidus  
C:Species: Archaeoglobus fulgidus  
C>Date: 05-Dec-1997 #sequence\_revision 05-Dec-1997 #text\_change 22-Oct-1999  
C:Accession: D69543  
R:Klenk, H.P.; Clayton, R.A.; Tomb, J.F.; White, O.; Nelson, K.E.; Ketchum, K.A.; Dodson  
., Fleischmann, R.D.; Quackenbush, J.; Lee, N.H.; Sutton, G.G.; Gill, S.; Kirkness, E.F.  
Glodek, A.; Zhou, L.; Overbeek, R.; Gocayne, J.D.; Weidman, J.F.; McDonald, L.  
Nature 390, 364-370, 1997  
A:Authors: Overbeek, T.; Cotton, M.D.; Spriggs, T.; Artlich, P.; Kaine, B.P.; Sykes, S.  
Smith, H.O.; Woese, C.R.; Venter, J.C.  
A:Title: The complete genome sequence of the hyperthermophilic, sulfate-reducing archaea  
A:Reference number: A69250; MUID:98049343; PMID:9389475  
A:Accession: D69543  
A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-112 <KLE>  
A:Cross-references: GB:AE001114; GB:AE000782; NID:G2689437; PIDN:AA891315.1; PID:G265074  
C:Superfamily: Archaeoglobus fulgidus hypothetical protein AF2348

Query Match 5.7%; Score 6; DB 2; Length 112;  
Best Local Similarity 100.0%; Pred. No. 91;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 95 DPSRKN 100  
|||  
Db 80 DPSRKN 85

RESULT 36  
G89134  
protein F25G6.8 [imported] - Caenorhabditis elegans  
C:Species: Caenorhabditis elegans  
C>Date: 10-May-2001 #sequence\_revision 10-May-2001 #text\_change 10-May-2001  
C:Accession: G89134  
R:anonymous, The C. elegans Sequencing Consortium.  
Science 282, 2012-2018, 1998  
A:Title: Genome sequence of the nematode C. elegans: a platform for investigating biolog  
A:Reference number: A75000; MUID:99069613; PMID:9851916  
A:Note: see websites genome.wustl.edu/gsc/C.elegans/ and www.sanger.ac.uk/Projects/C.ela  
A:Note: published extrata appeared in Science 283, 35, 1999; Science 283, 2103, 1999; and  
A:Accession: G89134  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-116 <STO>  
A:Cross-references: GB:chr\_V; PIDN:AAC25801.1; PID:G2384838; GSPDB:GN00023; CBSP:F25G6.8  
C:Genetics:  
A:Gene: F25G6.8  
A:Map position: 5

Query Match 5.7%; Score 6; DB 2; Length 116;  
Best Local Similarity 100.0%; Pred. No. 94;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 55 AQMVNL 60  
|||  
Db 93 AQMVNL 98

RESULT 37  
G69532  
conserved hypothetical protein AF2263 - Archaeoglobus fulgidus  
C:Species: Archaeoglobus fulgidus  
C>Date: 05-Dec-1997 #sequence\_revision 05-Dec-1997 #text\_change 24-Nov-1999  
C:Accession: G69532  
R:Klenk, H.P.; Clayton, R.A.; Tomb, J.F.; White, O.; Nelson, K.E.; Ketchum, K.A.; Dodso  
., Fleischmann, R.D.; Quackenbush, J.; Lee, N.H.; Sutton, G.G.; Gill, S.; Kirkness, E.F.  
Glodek, A.; Zhou, L.; Overbeek, R.; Gocayne, J.D.; Weidman, J.F.; McDonald, L.  
Nature 390, 364-370, 1997  
A:Authors: Overbeek, T.; Cotton, M.D.; Spriggs, T.; Artlich, P.; Kaine, B.P.; Sykes, S.  
Smith, H.O.; Woese, C.R.; Venter, J.C.  
A:Title: The complete genome sequence of the hyperthermophilic, sulfate-reducing archae  
A:Reference number: A69250; MUID:98049343; PMID:9389475  
A:Accession: G69532  
A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-122 <KLE>  
A:Cross-references: GB:AE000948; GB:AE000782; NID:G2689271; PIDN:AA88987.1; PID:G26482  
C:Superfamily: Archaeoglobus fulgidus conserved hypothetical protein AF2263

Query Match 5.7%; Score 6; DB 2; Length 122;  
Best Local Similarity 100.0%; Pred. No. 98;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 86 GYGRDD 91  
|||  
Db 28 GYGRDD 33

RESULT 38  
F97868  
hypothetical protein RCJ350 [imported] - Rickettsia conorii (strain Malish 7)  
C:Species: Rickettsia conorii  
C>Date: 30-Sep-2001 #sequence\_revision 30-Sep-2001 #text\_change 30-Sep-2001  
C:Accession: F97868  
R:Osata, H.; Audic, S.; Renesto-Audiffren, P.; Fournier, P.E.; Barbe, V.; Samson, D.; R  
Science 293, 2093-2098, 2001  
A:Title: Mechanisms of Evolution in Rickettsia conorii and Rickettsia prowazekii.  
A:Reference number: A97700; MUID:21442074; PMID:11557893  
A:Accession: F97868  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-131 <KUR>  
A:Cross-references: GB:AE006914; PIDN:AAU03888.1; PID:G15620494; GSPDB:GN00173  
C:Genetics:  
A:Gene: RCJ350

Query Match 5.7%; Score 6; DB 2; Length 131;  
Best Local Similarity 100.0%; Pred. No. 16+02;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 66 VISLPQ 71  
|||  
Db 84 VISLPQ 89

RESULT 39  
F69523  
conserved hypothetical protein AF2190 - Archaeoglobus fulgidus  
C:Species: Archaeoglobus fulgidus  
C>Date: 05-Dec-1997 #sequence\_revision 05-Dec-1997 #text\_change 22-Oct-1999  
C:Accession: F69523  
R:Klenk, H.P.; Clayton, R.A.; Tomb, J.F.; White, O.; Nelson, K.E.; Ketchum, K.A.; Dodso  
., Fleischmann, R.D.; Quackenbush, J.; Lee, N.H.; Sutton, G.G.; Gill, S.; Kirkness, E.F.  
Glodek, A.; Zhou, L.; Overbeek, R.; Gocayne, J.D.; Weidman, J.F.; McDonald, L.  
Nature 390, 364-370, 1997

A:Authors: Uterback, T.; Cotton, M.D.; Spriggs, T.; Artach, P.; Kaine, B.P.; Sykes, S.; Smith, H.O.; Woese, C.R.; Venter, J.C.  
 A:Title: The complete genome sequence of the hyperthermophilic, sulfate-reducing archaeon Pyrococcus furiosus  
 A:Reference number: A69250; MUID:98049343; PMID:9389475  
 A:Accession: F69523  
 A>Status: preliminary; nucleic acid sequence not shown; translation not shown  
 A:Molecule type: DNA  
 A:Residues: 1-136 <KLE>  
 A:Cross-references: GB:AE000954; GB:AE000782; NID:g2689277; PID:AA89064.1; PID:g264833

Query Match 5.7%; Score 6; DB 2; Length 136;  
 Best Local Similarity 100.0%; Pred. No. 1.1e+02;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 3 FTLRN 8  
 |||||  
 Db 74 FTLRN 79

## RESULT 40

R3HUI5

ribosomal protein S15, cytosolic [validated] - human

N:Alternate names: insulinoma protein rig

C:Species: Homo sapiens (man)

C:Date: 31-Mar-1992 #sequence\_revision 31-Mar-1992 #text\_change 08-Dec-2000

C:Accession: A35908; A33885; S68923

R:Shiga, K.; Yamamoto, H.; Okamoto, H.

Proc. Natl. Acad. Sci. U.S.A. 87, 3594-3598, 1990

A:Title: Isolation and characterization of the human homologue of rig and its pseudogene

A:Reference number: A35908; MUID:90239060; PMID:2159154

A:Accession: A35908

A:Molecule type: DNA

A:Residues: 1-145 &lt;SHI&gt;

A:Cross-references: EMBL:M32405; NID:g337415; PID:AA36568.1; PID:g337416

R:Imoue, C.; Shiga, K.; Takasawa, S.; Kitegawa, M.; Yamamoto, H.; Okamoto, H.

Proc. Natl. Acad. Sci. U.S.A. 84, 6659-6662, 1987

A:Title: Evolutionary conservation of the insulinoma gene rig and its possible function.

A:Reference number: A94183; MUID:88016150; PMID:2821540

A:Accession: A33885

A:Molecule type: mRNA

A:Residues: 1-145 &lt;INO&gt;

A:Cross-references: GB:J02984; NID:g184553; PID:AA36036.1; PID:g306898

R:Vladimirov, S.N.; Ivanov, A.V.; Karpova, G.G.; Musolyamov, A.K.; Egorov, T.A.; Thiede, Eur. J. Biochem. 239, 144-149, 1996

A:Title: Characterization of the human small-ribosomal-subunit proteins by N-terminal an

A:Reference number: S68911; MUID:96305378; PMID:8706699

A:Accession: S68923

A:Molecule type: protein

A:Residues: 101-113, 'X', 115-116 &lt;VLA&gt;

C:Genetics:

A:Gene: GDB:RPS15

A:Cross-references: GDB:132222; OMIM:180535

A:Map position: 5q31-5q33

A:Introns: 1/3; 30/2; 108/3

C:Superfamily: Escherichia coli ribosomal protein S19

F:2-145/Product: ribosomal protein S15 #status experimental &lt;MAT&gt;

F:61-68/Region: nuclear location signal

F:2/Modified site: blocked amino end (Ala) (in mature form) (probably acetylated) #stratu

Query Match 5.7%; Score 6; DB 1; Length 145;  
 Best Local Similarity 100.0%; Pred. No. 1.1e+02;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 33 SARORR 38  
 |||||  
 Db 38 SARORR 43

Search completed: April 9, 2003, 13:15:58  
 Job time : 74.1803 secs



GenCore version 5.1.4 p5 4578  
Copyright (c) 1993 - 2003 CompuGen Ltd.

## OM protein - protein search, using sw model

Run on: April 9, 2003, 13:10:22 ; Search time 33.0164 Seconds  
(without alignments)  
94.463 Million cell updates/sec

Title: US-10-020-441-2

Perfect score: 106  
Sequence: 1 VSFLLRNGDQGIHQPSSG.....YGRDDNDPRSRKNGILKK 106

Scoring table: OLIGO  
Gapop 60.0 , Gapext 60.0

Searched: 262574 seqs, 29422922 residues

Word size : 0

Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Listing first 1000 summaries

## Database :

Issued Patents\_AA:\*  
1: /cgn2\_6/ptodata/1/1aa/5A\_COMB.pep:\*  
2: /cgn2\_6/ptodata/1/1aa/5B\_COMB.pep:\*  
3: /cgn2\_6/ptodata/1/1aa/6A\_COMB.pep:\*  
4: /cgn2\_6/ptodata/1/1aa/6B\_COMB.pep:\*  
5: /cgn2\_6/ptodata/1/1aa/PTOS\_COMB.pep:\*  
6: /cgn2\_6/ptodata/1/1aa/Backfiles1.pep:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	7	6.6	725	3	US-09-234-332-12
2	6	5.7	60	4	US-08-817-787-22
3	6	5.7	74	4	US-09-134-001C-4293
4	6	5.7	80	1	US-08-225-757B-14
5	6	5.7	80	1	US-08-011-398B-16
6	6	5.7	80	1	US-08-464-051-16
7	6	5.7	80	2	US-08-462-498-16
8	6	5.7	80	3	US-08-554-385-15
9	6	5.7	95	4	US-09-142-078-66
10	6	5.7	95	4	US-09-357-141-66
11	6	5.7	95	4	US-09-533-889-66
12	6	5.7	113	2	US-08-710-330A-1
13	6	5.7	113	2	US-08-710-330A-3
14	6	5.7	113	2	US-08-710-330A-4
15	6	5.7	113	2	US-08-710-330A-5
16	6	5.7	154	2	US-08-710-330A-11
17	6	5.7	154	4	US-08-817-787-18
18	6	5.7	168	4	US-09-512-342-20
19	6	5.7	170	4	US-09-199-637A-158
20	6	5.7	190	2	US-08-531-525-45
21	6	5.7	190	2	US-08-718-270A-45
22	6	5.7	215	4	US-09-170-769A-4
23	6	5.7	269	1	US-07-706-691G-1
24	6	5.7	269	1	US-07-706-691G-2
25	6	5.7	269	1	US-07-706-691G-3
26	6	5.7	269	1	US-07-706-691G-4
27	6	5.7	269	1	US-07-706-691G-5

28	6	5.7	269	1	US-07-706-691G-6	Sequence 6, Appli
29	6	5.7	269	1	US-07-706-691G-7	Sequence 7, Appli
30	6	5.7	269	1	US-07-706-691G-8	Sequence 8, Appli
31	6	5.7	269	1	US-07-706-691G-9	Sequence 9, Appli
32	6	5.7	269	1	US-07-706-691G-10	Sequence 10, Appli
33	6	5.7	269	1	US-07-706-691G-11	Sequence 11, Appli
34	6	5.7	269	1	US-07-706-691G-12	Sequence 12, Appli
35	6	5.7	269	1	US-07-706-691G-13	Sequence 13, Appli
36	6	5.7	269	1	US-07-706-691G-14	Sequence 14, Appli
37	6	5.7	269	1	US-07-706-691G-15	Sequence 15, Appli
38	6	5.7	269	1	US-07-706-691G-16	Sequence 16, Appli
39	6	5.7	269	1	US-07-706-691G-17	Sequence 17, Appli
40	6	5.7	269	1	US-07-706-691G-18	Sequence 18, Appli
41	6	5.7	269	1	US-07-706-691G-19	Sequence 19, Appli
42	6	5.7	269	1	US-07-706-691G-20	Sequence 20, Appli
43	6	5.7	269	1	US-07-706-691G-21	Sequence 21, Appli
44	6	5.7	269	1	US-07-706-691G-22	Sequence 22, Appli
45	6	5.7	269	1	US-07-706-691G-23	Sequence 23, Appli
46	6	5.7	269	1	US-07-706-691G-24	Sequence 24, Appli
47	6	5.7	269	1	US-07-706-691G-25	Sequence 25, Appli
48	6	5.7	269	1	US-07-706-691G-26	Sequence 26, Appli
49	6	5.7	269	1	US-07-706-691G-27	Sequence 27, Appli
50	6	5.7	269	1	US-07-706-691G-28	Sequence 28, Appli
51	6	5.7	269	1	US-07-706-691G-29	Sequence 29, Appli
52	6	5.7	269	1	US-07-706-691G-30	Sequence 30, Appli
53	6	5.7	269	1	US-07-706-691G-31	Sequence 31, Appli
54	6	5.7	269	1	US-07-706-691G-32	Sequence 32, Appli
55	6	5.7	269	1	US-07-706-691G-33	Sequence 33, Appli
56	6	5.7	269	1	US-07-706-691G-34	Sequence 34, Appli
57	6	5.7	269	1	US-07-706-691G-35	Sequence 35, Appli
58	6	5.7	269	1	US-07-706-691G-36	Sequence 36, Appli
59	6	5.7	269	1	US-07-706-691G-37	Sequence 37, Appli
60	6	5.7	269	1	US-07-706-691G-38	Sequence 38, Appli
61	6	5.7	269	1	US-07-706-691G-39	Sequence 39, Appli
62	6	5.7	269	1	US-07-706-691G-40	Sequence 40, Appli
63	6	5.7	269	1	US-07-706-691G-41	Sequence 41, Appli
64	6	5.7	269	1	US-07-706-691G-42	Sequence 42, Appli
65	6	5.7	269	1	US-07-706-691G-43	Sequence 43, Appli
66	6	5.7	269	1	US-07-706-691G-44	Sequence 44, Appli
67	6	5.7	269	1	US-07-706-691G-45	Sequence 45, Appli
68	6	5.7	269	1	US-07-706-691G-46	Sequence 46, Appli
69	6	5.7	269	1	US-07-706-691G-47	Sequence 47, Appli
70	6	5.7	269	1	US-07-706-691G-48	Sequence 48, Appli
71	6	5.7	269	1	US-07-706-691G-49	Sequence 49, Appli
72	6	5.7	269	1	US-07-706-691G-50	Sequence 50, Appli
73	6	5.7	269	1	US-07-706-691G-51	Sequence 51, Appli
74	6	5.7	269	1	US-07-706-691G-52	Sequence 52, Appli
75	6	5.7	269	1	US-08-254-021-1	Sequence 1, Appli
76	6	5.7	269	1	US-08-254-021-2	Sequence 2, Appli
77	6	5.7	269	1	US-08-254-021-3	Sequence 3, Appli
78	6	5.7	269	1	US-08-254-021-4	Sequence 4, Appli
79	6	5.7	269	1	US-08-254-021-5	Sequence 5, Appli
80	6	5.7	269	1	US-08-254-021-6	Sequence 6, Appli
81	6	5.7	269	1	US-08-254-021-7	Sequence 7, Appli
82	6	5.7	269	1	US-08-254-021-8	Sequence 8, Appli
83	6	5.7	269	1	US-08-254-021-9	Sequence 9, Appli
84	6	5.7	269	1	US-08-254-021-10	Sequence 10, Appli
85	6	5.7	269	1	US-08-254-021-11	Sequence 11, Appli
86	6	5.7	269	1	US-08-254-021-12	Sequence 12, Appli
87	6	5.7	269	1	US-08-254-021-13	Sequence 13, Appli
88	6	5.7	269	1	US-08-254-021-14	Sequence 14, Appli
89	6	5.7	269	1	US-08-254-021-15	Sequence 15, Appli
90	6	5.7	269	1	US-08-254-021-16	Sequence 16, Appli
91	6	5.7	269	1	US-08-254-021-17	Sequence 17, Appli
92	6	5.7	269	1	US-08-254-021-18	Sequence 18, Appli
93	6	5.7	269	1	US-08-254-021-19	Sequence 19, Appli
94	6	5.7	269	1	US-08-254-021-20	Sequence 20, Appli
95	6	5.7	269	1	US-08-254-021-21	Sequence 21, Appli
96	6	5.7	269	1	US-08-254-021-22	Sequence 22, Appli
97	6	5.7	269	1	US-08-254-021-23	Sequence 23, Appli
98	6	5.7	269	1	US-08-254-021-24	Sequence 24, Appli
99	6	5.7	269	1	US-08-254-021-25	Sequence 25, Appli
100	6	5.7	269	1	US-08-254-021-26	Sequence 26, Appli

101	6	5.7	269	1	US-08-254-021-27	Sequence 27, Appl 1	174	5	5.7	269	2	US-08-618-446-31	Sequence 31, Appl 1
102	6	5.7	269	1	US-08-254-021-28	Sequence 28, Appl 1	175	6	5.7	269	2	US-08-618-446-32	Sequence 32, Appl 1
103	6	5.7	269	1	US-08-254-021-29	Sequence 29, Appl 1	176	6	5.7	269	2	US-08-618-446-33	Sequence 33, Appl 1
104	6	5.7	269	1	US-08-254-021-30	Sequence 30, Appl 1	177	6	5.7	269	2	US-08-618-446-34	Sequence 34, Appl 1
105	6	5.7	269	1	US-08-254-021-31	Sequence 31, Appl 1	178	6	5.7	269	2	US-08-618-446-35	Sequence 35, Appl 1
106	6	5.7	269	1	US-08-254-021-32	Sequence 32, Appl 1	179	6	5.7	269	2	US-08-618-446-36	Sequence 36, Appl 1
107	6	5.7	269	1	US-08-254-021-33	Sequence 33, Appl 1	180	6	5.7	269	2	US-08-618-446-37	Sequence 37, Appl 1
108	6	5.7	269	1	US-08-254-021-34	Sequence 34, Appl 1	181	6	5.7	269	2	US-08-618-446-38	Sequence 38, Appl 1
109	6	5.7	269	1	US-08-254-021-35	Sequence 35, Appl 1	182	6	5.7	269	2	US-08-618-446-39	Sequence 39, Appl 1
110	6	5.7	269	1	US-08-254-021-36	Sequence 36, Appl 1	183	6	5.7	269	2	US-08-618-446-40	Sequence 40, Appl 1
111	6	5.7	269	1	US-08-254-021-37	Sequence 37, Appl 1	184	6	5.7	269	2	US-08-618-446-41	Sequence 41, Appl 1
112	6	5.7	269	1	US-08-254-021-38	Sequence 38, Appl 1	185	6	5.7	269	2	US-08-618-446-42	Sequence 42, Appl 1
113	6	5.7	269	1	US-08-254-021-39	Sequence 39, Appl 1	186	6	5.7	269	2	US-08-618-446-43	Sequence 43, Appl 1
114	6	5.7	269	1	US-08-254-021-40	Sequence 40, Appl 1	187	6	5.7	269	2	US-08-618-446-44	Sequence 44, Appl 1
115	6	5.7	269	1	US-08-254-021-41	Sequence 41, Appl 1	188	6	5.7	269	2	US-08-618-446-45	Sequence 45, Appl 1
116	6	5.7	269	1	US-08-254-021-42	Sequence 42, Appl 1	189	6	5.7	269	2	US-08-618-446-46	Sequence 46, Appl 1
117	6	5.7	269	1	US-08-254-021-43	Sequence 43, Appl 1	190	6	5.7	269	2	US-08-618-446-47	Sequence 47, Appl 1
118	6	5.7	269	1	US-08-254-021-44	Sequence 44, Appl 1	191	6	5.7	269	2	US-08-618-446-48	Sequence 48, Appl 1
119	6	5.7	269	1	US-08-254-021-45	Sequence 45, Appl 1	192	6	5.7	269	2	US-08-618-446-49	Sequence 49, Appl 1
120	6	5.7	269	1	US-08-254-021-46	Sequence 46, Appl 1	193	6	5.7	269	2	US-08-618-446-50	Sequence 50, Appl 1
121	6	5.7	269	1	US-08-254-021-47	Sequence 47, Appl 1	194	6	5.7	269	2	US-08-618-446-51	Sequence 51, Appl 1
122	6	5.7	269	1	US-08-254-021-48	Sequence 48, Appl 1	195	6	5.7	269	2	US-08-618-446-52	Sequence 52, Appl 1
123	6	5.7	269	1	US-08-254-021-49	Sequence 49, Appl 1	196	6	5.7	269	3	US-08-898-218-10	Sequence 10, Appl 1
124	6	5.7	269	1	US-08-254-021-50	Sequence 50, Appl 1	197	6	5.7	269	3	US-08-848-793-10	Sequence 10, Appl 1
125	6	5.7	269	1	US-08-254-021-51	Sequence 51, Appl 1	198	6	5.7	269	4	US-08-980-135-1	Sequence 1, Appl 1
126	6	5.7	269	1	US-08-254-021-52	Sequence 52, Appl 1	199	6	5.7	269	4	US-08-980-135-2	Sequence 2, Appl 1
127	6	5.7	269	1	US-08-431-387-5	Sequence 5, Appl 1	200	6	5.7	269	4	US-08-980-135-3	Sequence 3, Appl 1
128	6	5.7	269	1	US-08-431-387-6	Sequence 6, Appl 1	201	6	5.7	269	4	US-08-980-135-4	Sequence 4, Appl 1
129	6	5.7	269	1	US-08-323-677A-10	Sequence 10, Appl 1	202						

247	6	5.7	269	4	US-08-980-135-50	Sequence 50, Appl	320	6	5.7	269	4	US-09-585-798-50	Sequence 50, Appl
248	6	5.7	269	4	US-08-980-135-51	Sequence 51, Appl	321	6	5.7	269	4	US-09-585-798-51	Sequence 51, Appl
249	6	5.7	269	4	US-08-980-135-52	Sequence 52, Appl	322	6	5.7	269	4	US-09-585-798-52	Sequence 52, Appl
250	6	5.7	269	4	US-09-074-331-10	Sequence 10, Appl	323	6	5.7	269	4	US-09-417-359A-3	Sequence 3, Appl
251	6	5.7	269	4	US-09-074-331-11	Sequence 11, Appl	324	6	5.7	269	5	PCT-US95-01937-10	Sequence 10, Appl
252	6	5.7	269	4	US-09-074-331-12	Sequence 12, Appl	325	6	5.7	269	5	PCT-US95-01937-11	Sequence 11, Appl
253	6	5.7	269	4	US-09-074-331-13	Sequence 13, Appl	326	6	5.7	269	5	PCT-US95-01937-12	Sequence 12, Appl
254	6	5.7	269	4	US-09-074-331-14	Sequence 14, Appl	327	6	5.7	269	5	PCT-US95-01937-13	Sequence 13, Appl
255	6	5.7	269	4	US-09-074-331-15	Sequence 15, Appl	328	6	5.7	269	5	PCT-US95-01937-14	Sequence 14, Appl
256	6	5.7	269	4	US-09-074-331-16	Sequence 16, Appl	329	6	5.7	269	5	PCT-US95-01937-15	Sequence 15, Appl
257	6	5.7	269	4	US-09-074-331-17	Sequence 17, Appl	330	6	5.7	269	5	PCT-US95-01937-16	Sequence 16, Appl
258	6	5.7	269	4	US-09-074-331-18	Sequence 18, Appl	331	6	5.7	269	5	PCT-US95-01937-17	Sequence 17, Appl
259	6	5.7	269	4	US-09-074-331-22	Sequence 22, Appl	332	6	5.7	269	5	PCT-US95-01937-18	Sequence 18, Appl
260	6	5.7	269	4	US-09-074-331-23	Sequence 23, Appl	333	6	5.7	269	5	PCT-US95-01937-19	Sequence 19, Appl
261	6	5.7	269	4	US-09-074-331-24	Sequence 24, Appl	334	6	5.7	269	5	5217878-3	Patent No. 5217878
262	6	5.7	269	4	US-09-255-502-5	Sequence 5, Appl	335	6	5.7	275	1	US-08-431-387-1	Sequence 1, Appl
263	6	5.7	269	4	US-09-024-532-3	Sequence 3, Appl	336	6	5.7	275	2	US-08-750-406A-1	Sequence 1, Appl
264	6	5.7	269	4	US-08-269-050-4	Sequence 4, Appl	337	6	5.7	275	2	US-09-337-118-1	Sequence 1, Appl
265	6	5.7	269	4	US-08-090-207-1	Sequence 1, Appl	338	6	5.7	276	1	US-08-431-387-7	Sequence 7, Appl
266	6	5.7	269	4	US-08-090-207-2	Sequence 2, Appl	339	6	5.7	332	4	US-08-858-207A-319	Sequence 319, App
267	6	5.7	269	4	US-09-178-155-6	Sequence 6, Appl	340	6	5.7	349	1	US-07-661-378A-2	Sequence 2, Appl
268	6	5.7	269	4	US-09-445-270-5	Sequence 5, Appl	341	6	5.7	355	3	US-08-872-879-7	Sequence 7, Appl
269	6	5.7	269	4	US-09-467-536A-1	Sequence 1, Appl	342	6	5.7	377	1	US-07-772-087-2	Sequence 2, Appl
270	6	5.7	269	4	US-09-234-957-1	Sequence 1, Appl	343	6	5.7	380	1	US-07-918-318-2	Sequence 2, Appl
271	6	5.7	269	4	US-09-585-798-1	Sequence 1, Appl	344	6	5.7	380	2	US-08-413-724-2	Sequence 2, Appl
272	6	5.7	269	4	US-09-585-798-2	Sequence 2, Appl	345	6	5.7	380	3	US-09-049-667-2	Sequence 2, Appl
273	6	5.7	269	4	US-09-585-798-3	Sequence 3, Appl	346	6	5.7	380	4	US-08-853-994-2	Sequence 2, Appl
274	6	5.7	269	4	US-09-585-798-4	Sequence 4, Appl	347	6	5.7	380	4	US-08-716-293-1	Sequence 1, Appl
275	6	5.7	269	4	US-09-585-798-5	Sequence 5, Appl	348	6	5.7	380	4	US-08-269-050-2	Sequence 2, Appl
276	6	5.7	269	4	US-09-585-798-6	Sequence 6, Appl	349	6	5.7	380	4	US-09-445-270-7	Sequence 7, Appl
277	6	5.7	269	4	US-09-585-798-7	Sequence 7, Appl	350	6	5.7	380	4	US-09-445-270-9	Sequence 9, Appl
278	6	5.7	269	4	US-09-585-798-8	Sequence 8, Appl	351	6	5.7	380	6	5217878-2	Patent No. 5217878
279	6	5.7	269	4	US-09-585-798-9	Sequence 9, Appl	352	6	5.7	380	6	536611-2	Patent No. 536611
280	6	5.7	269	4	US-09-585-798-10	Sequence 10, Appl	353	6	5.7	392	2	US-08-904-031-1	Sequence 1, Appl
281	6	5.7	269	4	US-09-585-798-11	Sequence 11, Appl	354	6	5.7	393	2	US-08-904-031-3	Sequence 3, Appl
282	6	5.7	269	4	US-09-585-798-12	Sequence 12, Appl	355	6	5.7	394	1	US-08-663-113A-2	Sequence 2, Appl
283	6	5.7	269	4	US-09-585-798-13	Sequence 13, Appl	356	6	5.7	394	3	US-09-014-888-2	Sequence 2, Appl
284	6	5.7	269	4	US-09-585-798-14	Sequence 14, Appl	357	6	5.7	417	4	US-09-574-414A-62	Sequence 62, Appl
285	6	5.7	269	4	US-09-585-798-15	Sequence 15, Appl	358	6	5.7	422	3	US-08-872-879-8	Sequence 8, Appl
286	6	5.7	269	4	US-09-585-798-16	Sequence 16, Appl	359	6	5.7	460	4	US-09-134-001C-532	Sequence 532, Ap
287	6	5.7	269	4	US-09-585-798-17	Sequence 17, Appl	360	6	5.7	466	4	US-09-134-001C-3526	Sequence 3526, Ap
288	6	5.7	269	4	US-09-585-798-18	Sequence 18, Appl	361	6	5.7	498	2	US-08-660-963-12	Sequence 12, Appl
289	6	5.7	269	4	US-09-585-798-19	Sequence 19, Appl	362	6	5.7	543	2	US-08-823-516-143	Sequence 143, App
290	6	5.7	269	4	US-09-585-798-20	Sequence 20, Appl	363	6	5.7	627	1	US-08-703-947-2	Sequence 2, Appl
291	6	5.7	269	4	US-09-585-798-21	Sequence 21, Appl	364	6	5.7	654	4	US-09-315-127-11	Sequence 11, Appl
292	6	5.7	269	4	US-09-585-798-22	Sequence 22, Appl	365	6	5.7	654	4	US-09-315-127-12	Sequence 12, Appl
293	6	5.7	269	4	US-09-585-798-23	Sequence 23, Appl	366	6	5.7	723	2	US-08-548-159-5	Sequence 5, Appl
294	6	5.7	269	4	US-09-585-798-24	Sequence 24, Appl	367	6	5.7	934	4	US-08-840-468A-19	Sequence 19, Appl
295	6	5.7	269	4	US-09-585-798-25	Sequence 25, Appl	368	6	5.7	934	4	US-09-696-188B-19	Sequence 19, Appl
296	6	5.7	269	4	US-09-585-798-26	Sequence 26, Appl	369	6	5.7	1001	4	US-08-884-669A-2	Sequence 2, Appl
297	6	5.7	269	4	US-09-585-798-27	Sequence 27, Appl	370	6	5.7	1183	4	US-09-134-001C-3530	Sequence 3530, Ap
298	6	5.7	269	4	US-09-585-798-28	Sequence 28, Appl	371	6	5.7	1200	3	US-08-840-006-5	Sequence 5, Appl
299	6	5.7	269	4	US-09-585-798-29	Sequence 29, Appl	372	6	5.7	1289	2	US-08-542-003-2	Sequence 2, Appl
300	6	5.7	269	4	US-09-585-798-30	Sequence 30, Appl	373	6	5.7	1289	2	US-08-332-760A-2	Sequence 2, Appl
301	6	5.7	269	4	US-09-585-798-31	Sequence 31, Appl	374	6	5.7	1289	4	US-09-236-449-2	Sequence 2, Appl
302	6	5.7	269	4	US-09-585-798-32	Sequence 32, Appl	375	6	5.7	2161	3	US-09-081-320-3	Sequence 3, Appl
303	6	5.7	269	4	US-09-585-798-33	Sequence 33, Appl	376	6	5.7	2161	4	US-09-574-414A-3	Sequence 3, Appl
304	6	5.7	269	4	US-09-585-798-34	Sequence 34, Appl	377	6	5.7	2161	4	US-09-707-780-3	Sequence 3, Appl
305	6	5.7	269	4	US-09-585-798-35	Sequence 35, Appl	378	6	5.7	2165	1	US-08-514-975B-2	Sequence 2, Appl
306	6	5.7	269	4	US-09-585-798-36	Sequence 36, Appl	379	6	5.7	2165	4	PCT-US95-12507-2	Sequence 2, Appl
307	6	5.7	269	4	US-09-585-798-37	Sequence 37, Appl	380	6	5.7	8	4	US-09-183-266A-32	Sequence 32, Appl
308	6	5.7	269	4	US-09-585-798-38	Sequence 38, Appl	381	5	4.7	9	1	US-08-787-547-66	Sequence 66, Appl
309	6	5.7	269	4	US-09-585-798-39	Sequence 39, Appl	382	5	4.7	9	5	PCT-US95-16415-3	Sequence 3, Appl
310	6	5.7	269	4	US-09-585-798-40	Sequence 40, Appl	383	5	4.7	10	3	US-08-159-339A-298	Sequence 298, App
311	6	5.7	269	4	US-09-585-798-41	Sequence 41, Appl	384	5	4.7	10	6	PCT-US95-16415-27	Sequence 27, Appl
312	6	5.7	269	4	US-09-585-798-42	Sequence 42, Appl	385	5	4.7	10	6	5252466-18	Patent No. 5252466
313	6	5.7	269	4	US-09-585-798-43	Sequence 43, Appl	386	5	4.7	12	4	US-08-871-732A-14	Sequence 14, Appl
314	6	5.7	269	4	US-09-585-798-44	Sequence 44, Appl	387	5	4.7	12	4	US-09-346-510B-14	Sequence 14, Appl
315	6	5.7	269	4	US-09-585-798-45	Sequence 45, Appl	388	5	4.7	12	5	PCT-US93-05701-16	Sequence 16, Appl
316	6	5.7	269	4	US-09-585-798-46	Sequence 46, Appl	389	5	4.7	15	1	US-08-268-251-9	Sequence 9, Appl
317	6	5.7	269	4	US-09-585-798-47	Sequence 47, Appl	390	5	4.7	15	5	PCT-US93-01112-9	Sequence 9, Appl
318	6	5.7	269	4	US-09-585-798-48	Sequence 48, Appl	391	5	4.7	16	1	US-08-307-724B-21	Sequence 21, Appl
319	6	5.7	269	4	US-09-585-798-49	Sequence 49, Appl	392	5	4.7	18	4	US-09-152-060-113	Sequence 113, App

393	5	4.7	18	5	PCT-US94-01234-2	Sequence 2, Appli	466	5	4.7	72	4	US-09-042-071-38	Sequence 38, Appli
394	5	4.7	19	2	US-08-374-560-7	Sequence 7, Appli	467	5	4.7	75	1	US-08-422-613-4	Sequence 4, Appli
395	5	4.7	20	1	US-08-786-748A-105	Sequence 105, App	468	5	4.7	75	1	US-09-288-143-205	Sequence 205, Appli
396	5	4.7	20	2	US-08-932-682-105	Sequence 105, App	469	5	4.7	79	4	US-09-134-001C-5135	Sequence 5135, Ap
397	5	4.7	20	4	US-08-122-458D-7	Sequence 7, Appli	470	5	4.7	84	5	PCT-US93-08386-17	Sequence 17, Appli
398	5	4.7	22	2	US-08-482-142-87	Sequence 87, Appli	471	5	4.7	87	2	US-08-562-114B-24	Sequence 24, Appli
399	5	4.7	22	2	US-08-478-572-87	Sequence 87, Appli	472	5	4.7	87	2	US-08-562-114B-25	Sequence 25, Appli
400	5	4.7	22	4	US-08-484-296-87	Sequence 87, Appli	473	5	4.7	87	2	US-08-729-594A-24	Sequence 24, Appli
401	5	4.7	23	1	US-08-268-251-39	Sequence 39, Appli	474	5	4.7	87	4	US-08-729-594A-25	Sequence 25, Appli
402	5	4.7	23	2	US-08-482-142-86	Sequence 86, Appli	475	5	4.7	87	4	US-08-937-993-24	Sequence 24, Appli
403	5	4.7	23	2	US-08-482-142-102	Sequence 102, App	476	5	4.7	87	4	US-08-937-993-25	Sequence 25, Appli
404	5	4.7	23	2	US-08-478-572-86	Sequence 86, Appli	477	5	4.7	87	4	US-09-101-468-39	Sequence 39, Appli
405	5	4.7	23	2	US-08-478-572-102	Sequence 102, App	478	5	4.7	91	2	US-08-273-146-19	Sequence 40, Appli
406	5	4.7	23	2	US-08-808-277A-21	Sequence 21, Appli	479	5	4.7	92	4	US-09-334-601-28	Sequence 28, Appli
407	5	4.7	23	2	US-08-808-277A-24	Sequence 24, Appli	480	5	4.7	98	2	US-08-825-556A-4	Sequence 4, Appli
408	5	4.7	23	4	US-08-484-296-86	Sequence 86, Appli	481	5	4.7	100	1	US-08-202-389-26	Sequence 26, Appli
409	5	4.7	23	4	US-08-484-296-102	Sequence 102, App	482	5	4.7	100	4	US-08-679-493A-146	Sequence 146, App
410	5	4.7	23	5	PCT-US93-01112-39	Sequence 39, Appli	483	5	4.7	105	1	US-08-474-177-14	Sequence 14, Appli
411	5	4.7	24	1	US-08-141-324-26	Sequence 26, Appli	484	5	4.7	105	1	US-08-487-033-14	Sequence 14, Appli
412	5	4.7	24	1	US-08-541-902-26	Sequence 26, Appli	485	5	4.7	105	1	US-08-340-203A-8	Sequence 8, Appli
413	5	4.7	24	4	US-09-227-357-493	Sequence 493, App	486	5	4.7	105	1	US-08-480-810-14	Sequence 14, Appli
414	5	4.7	25	2	US-08-482-142-100	Sequence 100, App	487	5	4.7	105	2	US-08-508-735-14	Sequence 14, Appli
415	5	4.7	25	2	US-08-478-572-100	Sequence 100, App	488	5	4.7	105	2	US-08-452-427-8	Sequence 8, Appli
416	5	4.7	25	4	US-08-484-296-100	Sequence 100, App	489	5	4.7	105	2	US-08-848-251-14	Sequence 14, Appli
417	5	4.7	26	4	US-07-757-022B-16	Sequence 16, Appli	490	5	4.7	105	2	US-08-486-047-14	Sequence 14, Appli
418	5	4.7	27	1	US-07-971-096-7	Sequence 7, Appli	491	5	4.7	105	3	US-09-120-130-14	Sequence 14, Appli
419	5	4.7	27	1	US-07-956-700B-35	Sequence 35, Appli	492	5	4.7	105	3	US-09-115-252-14	Sequence 14, Appli
420	5	4.7	27	1	US-08-175-096-7	Sequence 7, Appli	493	5	4.7	105	3	US-08-986-515-14	Sequence 14, Appli
421	5	4.7	27	1	US-08-475-537-35	Sequence 35, Appli	494	5	4.7	105	3	US-09-085-407-8	Sequence 8, Appli
422	5	4.7	27	1	US-08-485-607-35	Sequence 35, Appli	495	5	4.7	105	4	US-09-120-128-14	Sequence 14, Appli
423	5	4.7	27	2	US-08-475-879-35	Sequence 35, Appli	496	5	4.7	105	4	US-09-120-129-14	Sequence 14, Appli
424	5	4.7	27	4	US-08-936-165A-335	Sequence 335, App	497	5	4.7	105	4	US-09-201-139-14	Sequence 14, Appli
425	5	4.7	27	4	US-09-433-043B-35	Sequence 35, Appli	498	5	4.7	105	4	US-09-120-131-14	Sequence 14, Appli
426	5	4.7	28	2	US-08-724-354D-6	Sequence 6, Appli	499	5	4.7	105	4	US-09-288-143-39	Sequence 39, Appli
427	5	4.7	28	3	US-09-270-984A-6	Sequence 6, Appli	500	5	4.7	106	4	US-09-149-476-341	Sequence 341, App
428	5	4.7	29	2	US-08-482-142-101	Sequence 101, App	501	5	4.7	107	1	US-08-276-852-103	Sequence 103, App
429	5	4.7	29	2	US-08-478-572-101	Sequence 101, App	502	5	4.7	107	1	US-08-899-575-103	Sequence 103, App
430	5	4.7	29	4	US-08-484-296-101	Sequence 101, App	503	5	4.7	107	1	US-08-899-575-103	Sequence 103, App
431	5	4.7	31	4	US-09-162-934-9	Sequence 9, Appli	504	5	4.7	107	5	PCT-US95-08743-103	Sequence 103, App
432	5	4.7	32	2	US-08-452-075-2	Sequence 2, Appli	505	5	4.7	108	1	US-08-276-852-109	Sequence 109, App
433	5	4.7	32	4	US-09-011-762-6	Sequence 6, Appli	506	5	4.7	108	1	US-08-899-575-109	Sequence 109, App
434	5	4.7	32	4	US-08-199-776-6	Sequence 6, Appli	507	5	4.7	108	1	US-08-899-575-109	Sequence 109, App
435	5	4.7	34	3	US-08-663-731-6	Sequence 6, Appli	508	5	4.7	108	2	US-08-378-939-16	Sequence 16, Appli
436	5	4.7	34	3	US-08-879-338-6	Sequence 6, Appli	509	5	4.7	108	2	US-08-378-939-20	Sequence 20, Appli
437	5	4.7	34	5	US-09-293-238B-6	Sequence 6, Appli	510	5	4.7	108	5	US-08-974-899-1	Sequence 1, Appli
438	5	4.7	34	5	PCT-US95-02044-6	Sequence 6, Appli	511	5	4.7	108	5	PCT-US95-08743-109	Sequence 109, App
439	5	4.7	37	2	US-08-726-306A-147	Sequence 147, App	512	5	4.7	109	1	US-08-448-196A-1	Sequence 1, Appli
440	5	4.7	37	4	US-09-149-476-741	Sequence 741, App	513	5	4.7	110	1	US-08-017-570-2	Sequence 2, Appli
441	5	4.7	39	2	US-08-023-980B-36	Sequence 36, Appli	514	5	4.7	110	1	US-08-422-613-2	Sequence 2, Appli
442	5	4.7	39	2	US-08-486-953A-31	Sequence 31, Appli	515	5	4.7	110	1	US-08-471-426-2	Sequence 2, Appli
443	5	4.7	40	4	US-09-065-383-30	Sequence 30, Appli	516	5	4.7	110	4	US-09-672-609-13	Sequence 13, Appli
444	5	4.7	42	2	US-08-480-229C-25	Sequence 25, Appli	517	5	4.7	110	4	US-09-672-609-15	Sequence 15, Appli
445	5	4.7	42	2	US-08-659-235C-25	Sequence 25, Appli	518	5	4.7	110	4	US-09-672-609-16	Sequence 16, Appli
446	5	4.7	42	2	US-08-488-161-71	Sequence 71, Appli	519	5	4.7	110	4	US-09-672-609-17	Sequence 17, Appli
447	5	4.7	42	3	US-09-273-685-71	Sequence 71, Appli	520	5	4.7	110	4	US-09-025-403A-13	Sequence 13, Appli
448	5	4.7	42	4	US-09-106-568E-101	Sequence 101, App	521	5	4.7	110	4	US-09-025-403A-15	Sequence 15, Appli
449	5	4.7	42	5	PCT-US95-11934-71	Sequence 71, Appli	522	5	4.7	110	4	US-09-025-403A-16	Sequence 16, Appli
450	5	4.7	49	2	US-08-284-446-9	Sequence 9, Appli	523	5	4.7	110	4	US-09-025-403A-17	Sequence 17, Appli
451	5	4.7	50	4	US-09-081-975-5	Sequence 5, Appli	524	5	4.7	110	5	PCT-US94-01709-2	Sequence 2, Appli
452	5	4.7	52	4	US-09-336-536-69	Sequence 69, Appli	525	5	4.7	113	4	US-09-378-088A-118	Sequence 118, App
453	5	4.7	54	4	US-09-306-446C-13	Sequence 13, Appli	526	5	4.7	119	2	US-08-785-065-7	Sequence 7, Appli
454	5	4.7	55	1	US-08-199-776-4	Sequence 4, Appli	527	5	4.7	119	4	US-09-151-412-7	Sequence 7, Appli
455	5	4.7	55	3	US-08-663-731-4	Sequence 4, Appli	528	5	4.7	122	4	US-09-134-001C-4987	Sequence 4987, Ap
456	5	4.7	55	3	US-08-879-338-4	Sequence 4, Appli	529	5	4.7	123	4	US-09-188-930-179	Sequence 179, App
457	5	4.7	55	4	US-09-293-238B-4	Sequence 4, Appli	530	5	4.7	124	4	US-09-042-071-39	Sequence 39, Appli
458	5	4.7	55	5	PCT-US95-02044-4	Sequence 4, Appli	531	5	4.7	124	4	US-09-485-316A-9	Sequence 9, Appli
459	5	4.7	60	4	US-09-314-268-143	Sequence 143, App	532	5	4.7	125	4	US-09-372-422A-6	Sequence 6, Appli
460	5	4.7	62	4	US-09-268-364-10	Sequence 10, Appli	533	5	4.7	126	4	US-08-751-359-12	Sequence 12, Appli
461	5	4.7	65	4	US-09-268-364-6	Sequence 6, Appli	534	5	4.7	126	4	US-08-907-146-12	Sequence 12, Appli
462	5	4.7	68	2	US-08-659-251-31	Sequence 31, Appli	535	5	4.7	129	3	US-08-911-853-31	Sequence 31, Appli
463	5	4.7	68	4	US-09-256-490-31	Sequence 31, Appli	536	5	4.7	129	4	US-09-479-409-21	Sequence 21, Appli
464	5	4.7	71	5	PCT-US96-11445-51	Sequence 51, Appli	537	5	4.7	129	4	US-09-479-453-21	Sequence 21, Appli
465	5	4.7	71	2	US-08-812-003-9	Sequence 9, Appli	538	5	4.7	129	4	US-09-325-932A-58	Sequence 58, Appli

539	5	4.7	131	2	US-08-525-864A-6	Sequence 6, Appli	612	5	4.7	167	4	US-08-961-083-138	Sequence 138, App
540	5	4.7	134	1	US-08-259-372A-6	Sequence 6, Appli	613	5	4.7	167	4	US-08-759-628-3	Sequence 3, Appli
541	5	4.7	134	1	US-08-468-671-6	Sequence 6, Appli	614	5	4.7	168	4	US-08-444-628-9	Sequence 9, Appli
542	5	4.7	134	2	US-08-562-114B-29	Sequence 29, Appli	615	5	4.7	168	4	US-08-357-820-9	Sequence 9, Appli
543	5	4.7	134	4	US-08-729-594A-29	Sequence 29, Appli	616	5	4.7	169	3	US-08-094-071-3	Sequence 3, Appli
544	5	4.7	134	4	US-08-937-993-29	Sequence 29, Appli	617	5	4.7	169	3	US-08-094-071-4	Sequence 4, Appli
545	5	4.7	137	4	US-09-456-830-7	Sequence 7, Appli	618	5	4.7	169	3	PCT-US92-00878-3	Sequence 3, Appli
546	5	4.7	137	4	US-09-456-830-15	Sequence 15, Appli	619	5	4.7	169	5	PCT-US92-00878-4	Sequence 4, Appli
547	5	4.7	137	4	US-09-456-830-19	Sequence 19, Appli	620	5	4.7	170	3	US-08-094-071-2	Sequence 2, Appli
548	5	4.7	137	4	US-09-456-830-37	Sequence 37, Appli	621	5	4.7	170	3	US-08-094-071-5	Sequence 5, Appli
549	5	4.7	137	4	US-09-456-830-41	Sequence 41, Appli	622	5	4.7	170	3	US-08-094-071-6	Sequence 6, Appli
550	5	4.7	137	4	US-09-002-285-7	Sequence 7, Appli	623	5	4.7	170	3	US-08-094-071-7	Sequence 7, Appli
551	5	4.7	137	4	US-09-002-285-15	Sequence 15, Appli	624	5	4.7	170	3	US-09-228-986-116	Sequence 116, App
552	5	4.7	137	4	US-09-002-285-19	Sequence 19, Appli	625	5	4.7	170	5	PCT-US92-00878-2	Sequence 2, Appli
553	5	4.7	137	4	US-09-002-285-37	Sequence 37, Appli	626	5	4.7	170	5	PCT-US92-00878-5	Sequence 5, Appli
554	5	4.7	137	4	US-09-002-285-41	Sequence 41, Appli	627	5	4.7	170	5	PCT-US92-00878-6	Sequence 6, Appli
555	5	4.7	139	3	US-08-753-007A-33	Sequence 33, Appli	628	5	4.7	170	5	PCT-US92-00878-7	Sequence 7, Appli
556	5	4.7	139	4	US-09-398-486-33	Sequence 33, Appli	629	5	4.7	171	1	US-08-303-270-1	Sequence 1, Appli
557	5	4.7	141	2	US-08-766-858A-46	Sequence 46, Appli	630	5	4.7	171	4	US-09-134-001C-3256	Sequence 3256, Ap
558	5	4.7	146	1	US-08-678-369-6	Sequence 6, Appli	631	5	4.7	172	4	US-09-134-001C-5649	Sequence 5649, Ap
559	5	4.7	146	2	US-09-047-243-6	Sequence 6, Appli	632	5	4.7	173	2	US-08-658-639-12	Sequence 12, Appli
560	5	4.7	146	4	US-08-914-375C-41	Sequence 41, Appli	633	5	4.7	173	4	US-08-944-604-12	Sequence 12, Appli
561	5	4.7	146	4	US-08-914-375C-43	Sequence 43, Appli	634	5	4.7	176	1	US-08-036-210-9	Sequence 9, Appli
562	5	4.7	146	4	US-09-325-932A-68	Sequence 68, Appli	635	5	4.7	176	2	US-08-639-251-6	Sequence 6, Appli
563	5	4.7	147	4	US-08-122-458D-9	Sequence 9, Appli	636	5	4.7	176	2	US-08-449-609-9	Sequence 9, Appli
564	5	4.7	148	1	US-08-154-915-4	Sequence 4, Appli	637	5	4.7	176	4	US-09-256-490-6	Sequence 6, Appli
565	5	4.7	148	3	US-08-384-106A-16	Sequence 16, Appli	638	5	4.7	176	5	PCT-US96-11445-6	Sequence 6, Appli
566	5	4.7	148	3	US-08-384-106A-24	Sequence 24, Appli	639	5	4.7	178	4	US-09-042-071-42	Sequence 42, Appli
567	5	4.7	148	3	US-08-513-974B-335	Sequence 335, App	640	5	4.7	181	3	US-08-753-007A-4	Sequence 4, Appli
568	5	4.7	148	4	US-08-540-650B-14	Sequence 14, Appli	641	5	4.7	181	4	US-09-338-496-4	Sequence 4, Appli
569	5	4.7	149	5	PCT-US93-09945-4	Sequence 4, Appli	642	5	4.7	181	4	US-08-858-207A-512	Sequence 512, Appl
570	5	4.7	149	5	US-08-581-148C-22	Sequence 22, Appli	643	5	4.7	184	4	US-09-552-322-11	Sequence 11, Appl
571	5	4.7	150	3	US-09-057-570-6	Sequence 6, Appli	644	5	4.7	185	4	US-09-562-737-115	Sequence 115, Appl
572	5	4.7	151	2	US-08-723-050-10	Sequence 10, Appli	645	5	4.7	188	4	US-09-333-627B-19	Sequence 19, Appl
573	5	4.7	152	1	US-07-644-372-2	Sequence 2, Appli	646	5	4.7	191	4	US-08-858-207A-410	Sequence 410, App
574	5	4.7	152	3	US-09-208-718-6	Sequence 6, Appli	647	5	4.7	193	2	US-08-861-549-4	Sequence 4, Appli
575	5	4.7	153	4	US-08-679-493A-201	Sequence 201, App	648	5	4.7	195	4	US-09-042-071-40	Sequence 40, Appli
576	5	4.7	153	4	US-08-679-493A-202	Sequence 202, App	649	5	4.7	195	4	US-08-849-6348-3	Sequence 3, Appli
577	5	4.7	154	4	US-08-871-732A-1	Sequence 1, Appli	650	5	4.7	198	4	US-09-335-332A-151	Sequence 151, App
578	5	4.7	154	4	US-09-346-510B-1	Sequence 1, Appli	651	5	4.7	202	4	US-08-821-994-78	Sequence 78, Appli
579	5	4.7	154	4	US-09-228-986-99	Sequence 99, Appli	652	5	4.7	202	4	US-08-858-207A-332	Sequence 332, App
580	5	4.7	154	1	US-08-474-177-2	Sequence 2, Appli	653	5	4.7	203	1	US-08-063-552-6	Sequence 6, Appli
581	5	4.7	156	1	US-08-487-813-2	Sequence 2, Appli	654	5	4.7	203	5	PCT-US93-05704-6	Sequence 6, Appli
582	5	4.7	156	1	US-08-480-810-2	Sequence 2, Appli	655	5	4.7	204	3	US-08-906-769-147	Sequence 147, App
583	5	4.7	156	2	US-08-508-735-2	Sequence 2, Appli	656	5	4.7	204	3	US-08-906-616-147	Sequence 147, App
584	5	4.7	156	2	US-08-627-610-2	Sequence 2, Appli	657	5	4.7	204	4	US-08-639-075A-147	Sequence 147, App
585	5	4.7	156	2	US-08-306-511A-2	Sequence 2, Appli	658	5	4.7	204	4	US-09-012-431-147	Sequence 147, App
586	5	4.7	156	2	US-08-893-274-2	Sequence 2, Appli	659	5	4.7	204	4	US-09-012-592-147	Sequence 147, App
587	5	4.7	156	2	US-08-848-251-2	Sequence 2, Appli	660	5	4.7	204	4	US-08-906-613-147	Sequence 147, App
588	5	4.7	156	2	US-08-486-047-2	Sequence 2, Appli	661	5	4.7	204	4	US-09-325-532A-145	Sequence 145, App
589	5	4.7	156	3	US-09-120-130-2	Sequence 2, Appli	662	5	4.7	205	2	US-08-861-549-1	Sequence 1, Appli
590	5	4.7	156	3	US-08-581-918A-2	Sequence 2, Appli	663	5	4.7	205	2	US-08-861-549-3	Sequence 3, Appli
591	5	4.7	156	3	US-09-115-252-2	Sequence 2, Appli	664	5	4.7	205	3	US-08-842-576-1	Sequence 1, Appli
592	5	4.7	156	3	US-08-986-515-2	Sequence 2, Appli	665	5	4.7	205	3	US-09-213-397-1	Sequence 1, Appli
593	5	4.7	156	4	US-09-120-128-2	Sequence 2, Appli	666	5	4.7	205	3	US-09-416-489-1	Sequence 1, Appli
594	5	4.7	156	4	US-09-120-129-2	Sequence 2, Appli	667	5	4.7	207	4	US-09-185-501B-14	Sequence 14, Appli
595	5	4.7	156	4	US-09-201-139-2	Sequence 2, Appli	668	5	4.7	210	2	US-08-948-616-5	Sequence 5, Appli
596	5	4.7	156	4	US-08-346-147B-2	Sequence 2, Appli	669	5	4.7	210	2	US-08-892-272-2	Sequence 2, Appli
597	5	4.7	156	4	US-09-120-111-2	Sequence 2, Appli	670	5	4.7	210	2	US-09-193-510-5	Sequence 5, Appli
598	5	4.7	156	4	US-08-822-936-2	Sequence 2, Appli	671	5	4.7	210	4	US-08-961-083-36	Sequence 36, Appli
599	5	4.7	156	4	US-08-910-722-2	Sequence 2, Appli	672	5	4.7	210	4	US-09-368-402-5	Sequence 5, Appli
600	5	4.7	156	4	US-08-497-214D-2	Sequence 2, Appli	673	5	4.7	210	4	US-09-306-446C-9	Sequence 9, Appli
601	5	4.7	156	4	US-09-134-001C-4260	Sequence 4260, Ap	674	5	4.7	210	4	US-09-306-446C-16	Sequence 16, Appli
602	5	4.7	156	4	US-09-457-568-38	Sequence 28, Appli	675	5	4.7	211	1	US-07-915-966C-18	Sequence 18, Appli
603	5	4.7	156	4	US-09-457-646-28	Sequence 28, Appli	676	5	4.7	211	2	US-08-771-182-18	Sequence 18, Appli
604	5	4.7	156	5	PCT-US95-04636-2	Sequence 2, Appli	677	5	4.7	211	2	US-08-853-194-18	Sequence 18, Appli
605	5	4.7	157	1	US-08-257-999-4	Sequence 4, Appli	678	5	4.7	213	2	US-08-716-284-2	Sequence 2, Appli
606	5	4.7	157	3	US-08-581-918A-12	Sequence 12, Appli	679	5	4.7	213	4	US-09-268-364-4	Sequence 4, Appli
607	5	4.7	157	4	US-08-346-147B-12	Sequence 12, Appli	680	5	4.7	213	4	US-09-291-170A-5	Sequence 5, Appli
608	5	4.7	157	4	US-08-497-214D-12	Sequence 12, Appli	681	5	4.7	213	4	US-09-724-684-5	Sequence 5, Appli
609	5	4.7	157	5	PCT-US96-05252-5	Sequence 5, Appli	682	5	4.7	214	4	US-08-861-774E-44	Sequence 44, Appli
610	5	4.7	166	4	US-09-068-195-14	Sequence 14, Appli	683	5	4.7	214	4	US-09-291-170A-7	Sequence 7, Appli
611	5	4.7	166	4	US-09-068-195-22	Sequence 22, Appli	684	5	4.7	214	4	US-09-724-684-7	Sequence 7, Appli

685	5	4.7	215	4	US-09-291-170A-6	Sequence 6, Appli	758	5	4.7	258	4	US-08-887-534A-18	Sequence 18, Appli
686	5	4.7	215	4	US-09-724-884-6	Sequence 6, Appli	759	5	4.7	258	4	US-09-954-314-12	Sequence 12, Appli
687	5	4.7	217	4	US-09-291-170A-4	Sequence 4, Appli	760	5	4.7	258	5	PCT-US91-05801-29	Sequence 29, Appli
688	5	4.7	217	4	US-09-724-884-4	Sequence 4, Appli	761	5	4.7	259	3	US-09-082-090-2	Sequence 2, Appli
689	5	4.7	217	4	US-07-757-022B-76	Sequence 76, Appli	762	5	4.7	260	2	US-08-675-921B-9	Sequence 9, Appli
690	5	4.7	218	4	US-09-081-975-7	Sequence 7, Appli	763	5	4.7	261	1	US-07-971-096-2	Sequence 2, Appli
691	5	4.7	219	2	US-08-470-261-2	Sequence 2, Appli	764	5	4.7	261	1	US-08-175-096-2	Sequence 2, Appli
692	5	4.7	219	2	US-08-816-605-2	Sequence 2, Appli	765	5	4.7	261	4	US-09-414-436-3	Sequence 3, Appli
693	5	4.7	219	3	US-08-916-989B-2	Sequence 2, Appli	766	5	4.7	262	4	US-08-961-083-98	Sequence 98, Appli
694	5	4.7	219	4	US-09-432-253-2	Sequence 2, Appli	767	5	4.7	262	1	US-08-152-922A-5	Sequence 5, Appli
695	5	4.7	219	5	PCT-US94-13187-2	Sequence 2, Appli	768	5	4.7	265	1	US-08-461-731-2	Sequence 2, Appli
696	5	4.7	224	4	US-09-113-750A-7	Sequence 7, Appli	769	5	4.7	265	2	US-09-002-072B-2	Sequence 2, Appli
697	5	4.7	225	4	US-09-302-769-8	Sequence 8, Appli	770	5	4.7	265	4	US-09-314-199-2	Sequence 2, Appli
698	5	4.7	226	4	US-09-268-364-2	Sequence 2, Appli	771	5	4.7	265	5	US-09-314-198-2	Sequence 2, Appli
699	5	4.7	227	2	US-08-471-717-2	Sequence 2, Appli	772	5	4.7	265	4	PCT-US94-10465-3	Sequence 3, Appli
700	5	4.7	228	4	US-09-134-001C-4083	Sequence 4083, Ap	773	5	4.7	266	2	US-08-235-480-7	Sequence 7, Appli
701	5	4.7	228	4	US-09-457-568-24	Sequence 24, Appli	774	5	4.7	266	2	US-09-097-759-4	Sequence 4, Appli
702	5	4.7	228	4	US-09-457-646-24	Sequence 24, Appli	775	5	4.7	266	4	US-09-118-445-7	Sequence 7, Appli
703	5	4.7	229	4	US-08-751-359-22	Sequence 22, Appli	776	5	4.7	267	2	US-08-934-481-4	Sequence 4, Appli
704	5	4.7	229	4	US-08-907-146-22	Sequence 22, Appli	777	5	4.7	267	2	US-08-667-809B-2	Sequence 2, Appli
705	5	4.7	230	4	US-09-625-188-18	Sequence 18, Appli	778	5	4.7	267	4	US-09-290-602-4	Sequence 4, Appli
706	5	4.7	231	2	US-08-622-354-5	Sequence 5, Appli	779	5	4.7	270	4	US-08-961-083-06	Sequence 206, Ap
707	5	4.7	233	1	US-07-890-422B-1	Sequence 1, Appli	780	5	4.7	270	4	US-08-983-035A-32	Sequence 32, Appli
708	5	4.7	234	4	US-09-149-476-478	Sequence 478, App	781	5	4.7	271	4	US-09-414-436-1	Sequence 1, Appli
709	5	4.7	237	1	US-08-750-532-18	Sequence 18, Appli	782	5	4.7	272	4	US-09-101-135-8	Sequence 8, Appli
710	5	4.7	237	4	US-09-457-568-20	Sequence 20, Appli	783	5	4.7	272	4	US-09-029-213B-27	Sequence 27, Appli
711	5	4.7	237	4	US-09-457-646-20	Sequence 20, Appli	784	5	4.7	272	4	US-09-134-001C-2864	Sequence 2864, Ap
712	5	4.7	239	4	US-08-213-419B-17	Sequence 17, Appli	785	5	4.7	272	5	PCT-US95-02455-4	Sequence 4, Appli
713	5	4.7	242	3	US-09-286-690-11	Sequence 11, Appli	786	5	4.7	273	1	US-08-201-697-17	Sequence 17, Appli
714	5	4.7	243	4	US-09-413-814-20	Sequence 20, Appli	787	5	4.7	277	4	US-09-134-001C-5291	Sequence 5291, Ap
715	5	4.7	245	2	US-08-685-992-26	Sequence 26, Appli	788	5	4.7	278	3	US-08-522-813-4	Sequence 4, Appli
716	5	4.7	245	2	US-09-144-925-26	Sequence 26, Appli	789	5	4.7	278	3	US-09-260-283-2	Sequence 2, Appli
717	5	4.7	245	4	US-09-369-364A-11	Sequence 11, Appli	790	5	4.7	279	3	US-09-286-690-9	Sequence 9, Appli
718	5	4.7	247	1	US-08-241-766-13	Sequence 13, Appli	791	5	4.7	280	4	US-09-383-886-18	Sequence 18, Appli
719	5	4.7	247	4	US-09-199-637A-363	Sequence 363, App	792	5	4.7	284	1	US-08-061-889-2	Sequence 2, Appli
720	5	4.7	248	2	US-08-735-559-1	Sequence 1, Appli	793	5	4.7	284	1	US-08-462-611-2	Sequence 2, Appli
721	5	4.7	248	3	US-09-210-474-1	Sequence 1, Appli	794	5	4.7	284	4	US-08-623-428D-2	Sequence 2, Appli
722	5	4.7	248	4	US-09-539-774-1	Sequence 1, Appli	795	5	4.7	284	4	US-09-465-558-54	Sequence 54, Appli
723	5	4.7	249	3	US-08-486-099-102	Sequence 102, App	796	5	4.7	284	4	US-09-395-674B-4	Sequence 4, Appli
724	5	4.7	249	3	US-08-484-223B-102	Sequence 102, App	797	5	4.7	284	5	PCT-US94-05378-2	Sequence 2, Appli
725	5	4.7	249	3	US-08-919-597-102	Sequence 102, App	798	5	4.7	286	4	US-09-134-001C-4602	Sequence 4602, Ap
726	5	4.7	249	3	US-08-475-668A-102	Sequence 102, App	799	5	4.7	287	4	US-08-936-107A-12	Sequence 12, Appli
727	5	4.7	249	3	US-08-485-551A-102	Sequence 102, App	800	5	4.7	288	4	US-09-095-758-8	Sequence 8, Appli
728	5	4.7	249	3	US-08-471-913A-102	Sequence 102, App	801	5	4.7	288	4	US-09-422-968-8	Sequence 8, Appli
729	5	4.7	249	4	US-08-485-264A-102	Sequence 102, App	802	5	4.7	288	4	US-09-372-448A-2	Sequence 2, Appli
730	5	4.7	249	4	US-08-474-349A-102	Sequence 102, App	803	5	4.7	288	4	US-09-708-015A-8	Sequence 8, Appli
731	5	4.7	251	2	US-08-685-992-15	Sequence 15, Appli	804	5	4.7	288	4	US-09-091-952A-17	Sequence 17, Appli
732	5	4.7	251	2	US-09-144-925-15	Sequence 15, Appli	805	5	4.7	289	4	US-09-372-422A-2	Sequence 2, Appli
733	5	4.7	252	4	US-09-457-568-22	Sequence 22, Appli	806	5	4.7	292	4	US-09-372-422A-4	Sequence 4, Appli
734	5	4.7	252	4	US-09-457-646-22	Sequence 22, Appli	807	5	4.7	293	2	US-08-284-465-4	Sequence 4, Appli
735	5	4.7	253	4	US-08-983-035A-10	Sequence 30, Appli	808	5	4.7	293	4	US-09-134-001C-3323	Sequence 3323, Ap
736	5	4.7	255	1	US-08-236-918A-8	Sequence 8, Appli	809	5	4.7	294	2	US-08-284-465-3	Sequence 3, Appli
737	5	4.7	255	1	US-08-622-353-8	Sequence 8, Appli	810	5	4.7	294	2	US-08-284-465-8	Sequence 8, Appli
738	5	4.7	255	2	US-08-622-352A-10	Sequence 10, Appli	811	5	4.7	294	4	US-09-103-664A-5	Sequence 5, Appli
739	5	4.7	255	2	US-08-816-605-9	Sequence 9, Appli	812	5	4.7	296	4	US-09-134-001C-4990	Sequence 4990, Ap
740	5	4.7	255	3	US-08-826-390-10	Sequence 10, Appli	813	5	4.7	299	2	US-09-097-759-3	Sequence 3, Appli
741	5	4.7	255	4	US-09-006-353A-11	Sequence 11, Appli	814	5	4.7	300	4	US-07-640-029-5	Sequence 5, Appli
742	5	4.7	255	4	US-08-815-225-4	Sequence 4, Appli	815	5	4.7	300	4	US-08-439-992A-5	Sequence 5, Appli
743	5	4.7	255	4	US-09-007-097-2	Sequence 2, Appli	816	5	4.7	302	1	US-07-640-029-6	Sequence 6, Appli
744	5	4.7	255	4	US-09-227-357-488	Sequence 488, App	817	5	4.7	302	1	US-07-921-807B-7	Sequence 7, Appli
745	5	4.7	255	4	US-09-150-864A-8	Sequence 8, Appli	818	5	4.7	302	1	US-07-921-807B-8	Sequence 8, Appli
746	5	4.7	255	4	US-09-573-986-11	Sequence 11, Appli	819	5	4.7	302	1	US-08-441-944A-7	Sequence 7, Appli
747	5	4.7	255	5	PCT-US96-03965-8	Sequence 8, Appli	820	5	4.7	302	1	US-08-441-944A-8	Sequence 8, Appli
748	5	4.7	256	2	US-08-286-819A-41	Sequence 41, Appli	821	5	4.7	302	4	US-08-439-992A-6	Sequence 6, Appli
749	5	4.7	256	2	US-08-980-357-41	Sequence 41, Appli	822	5	4.7	303	2	US-08-846-762-20	Sequence 20, Appli
750	5	4.7	257	2	US-07-857-224B-41	Sequence 41, Appli	823	5	4.7	303	2	US-08-846-762-21	Sequence 21, Appli
751	5	4.7	257	2	US-08-467-265-16	Sequence 16, Appli	824	5	4.7	306	2	US-09-097-759-2	Sequence 2, Appli
752	5	4.7	257	4	US-08-467-265-16	Sequence 16, Appli	825	5	4.7	308	4	US-09-149-476-007	Sequence 607, Appli
753	5	4.7	257	4	US-09-407-891-16	Sequence 16, Appli	826	5	4.7	312	1	US-08-425-061-18	Sequence 18, Appli
754	5	4.7	257	4	US-09-134-001C-3562	Sequence 3562, Ap	827	5	4.7	312	2	US-08-825-886-18	Sequence 18, Appli
755	5	4.7	258	1	US-07-721-761A-29	Sequence 29, Appli	828	5	4.7	312	4	US-09-134-001C-5583	Sequence 5583, Ap
756	5	4.7	258	1	US-07-978-687-29	Sequence 29, Appli	829	5	4.7	313	2	US-08-463-081B-26	Sequence 26, Appli
757	5	4.7	258	4	US-09-504-358-12	Sequence 12, Appli	830	5	4.7	313	2	US-08-461-379A-26	Sequence 26, Appli

831	5	4.7	313	2	US-08-462-390B-26	Sequence 26, Appl	904	5	4.7	342	4	US-08-867-611-37	Sequence 37, Appl
832	5	4.7	313	3	US-08-463-074B-26	Sequence 26, Appl	905	5	4.7	342	4	US-09-134-001C-5198	Sequence 5198, Ap
833	5	4.7	313	3	US-08-465-585C-26	Sequence 26, Appl	906	5	4.7	343	3	US-08-978-741-C-17	Sequence 17, Appl
834	5	4.7	313	3	US-08-652-446-26	Sequence 26, Appl	907	5	4.7	343	3	US-09-120-365-11	Sequence 71, Appl
835	5	4.7	313	4	US-09-237-543-7	Sequence 7, Appl	908	5	4.7	343	4	US-09-515-039-71	Sequence 71, Appl
836	5	4.7	313	4	US-09-237-543-8	Sequence 8, Appl	909	5	4.7	343	4	US-09-333-729A-13	Sequence 13, Appl
837	5	4.7	313	4	US-09-237-543-9	Sequence 9, Appl	910	5	4.7	344	1	US-08-453-924-10	Sequence 10, Appl
838	5	4.7	313	4	US-09-644-450-7	Sequence 7, Appl	911	5	4.7	344	4	US-08-867-611-38	Sequence 38, Appl
839	5	4.7	313	4	US-09-644-450-8	Sequence 8, Appl	912	5	4.7	344	1	US-09-199-637A-108	Sequence 108, App
840	5	4.7	313	4	US-09-644-450-9	Sequence 9, Appl	913	5	4.7	345	1	US-08-031-148-4	Sequence 4, Appl
841	5	4.7	315	1	US-08-253-155A-34	Sequence 34, Appl	914	5	4.7	345	3	US-08-415-838-4	Sequence 4, Appl
842	5	4.7	316	4	US-08-723-594A-36	Sequence 36, Appl	915	5	4.7	345	4	US-09-194-905-12	Sequence 12, Appl
843	5	4.7	316	4	US-08-937-993-36	Sequence 36, Appl	916	5	4.7	345	4	US-09-205-169-4	Sequence 4, Appl
844	5	4.7	317	2	US-08-977-847-3	Sequence 3, Appl	917	5	4.7	348	1	US-08-118-270-13	Sequence 1, Appl
845	5	4.7	317	2	US-09-195-021-3	Sequence 3, Appl	918	5	4.7	348	1	US-08-459-489-10	Sequence 10, Appl
846	5	4.7	317	3	US-09-109-205-18	Sequence 18, Appl	919	5	4.7	348	1	US-08-458-686-10	Sequence 10, Appl
847	5	4.7	317	3	US-08-605-284B-17	Sequence 17, Appl	920	5	4.7	348	1	US-07-843-350C-10	Sequence 10, Appl
848	5	4.7	317	4	US-08-940-424-4	Sequence 4, Appl	921	5	4.7	348	1	US-09-457-568-14	Sequence 14, Appl
849	5	4.7	317	4	US-08-940-424-5	Sequence 5, Appl	922	5	4.7	348	4	US-09-457-568-14	Sequence 14, Appl
850	5	4.7	317	4	US-08-729-594A-37	Sequence 37, Appl	923	5	4.7	348	4	US-09-325-932A-160	Sequence 160, App
851	5	4.7	318	1	US-08-937-993-37	Sequence 37, Appl	924	5	4.7	348	5	PCT-US93-08528-13	Sequence 13, Appl
852	5	4.7	318	1	US-08-547-182-2	Sequence 2, Appl	925	5	4.7	350	4	US-08-637-670-39	Sequence 39, Appl
853	5	4.7	318	2	US-08-278-042C-3	Sequence 3, Appl	926	5	4.7	351	1	US-08-470-202-46	Sequence 46, Appl
854	5	4.7	318	4	US-08-723-594A-38	Sequence 38, Appl	927	5	4.7	351	1	US-08-471-770-46	Sequence 46, Appl
855	5	4.7	318	4	US-08-937-993-38	Sequence 38, Appl	928	5	4.7	351	2	US-08-468-059-46	Sequence 46, Appl
856	5	4.7	319	4	US-08-983-035A-42	Sequence 42, Appl	929	5	4.7	351	4	US-09-109-516-46	Sequence 46, Appl
857	5	4.7	319	4	US-09-393-627B-4	Sequence 4, Appl	930	5	4.7	352	4	US-08-867-611-39	Sequence 39, Appl
858	5	4.7	320	3	US-08-852-730-9	Sequence 9, Appl	931	5	4.7	353	1	US-08-034-650-11	Sequence 11, Appl
859	5	4.7	321	1	US-07-945-288-6	Sequence 6, Appl	932	5	4.7	353	1	US-08-449-015-11	Sequence 11, Appl
860	5	4.7	321	1	US-08-462-831-6	Sequence 6, Appl	933	5	4.7	353	1	US-08-983-035A-40	Sequence 40, Appl
861	5	4.7	321	1	US-08-461-809-6	Sequence 6, Appl	934	5	4.7	355	1	US-08-118-270-11	Sequence 11, Appl
862	5	4.7	321	1	US-08-461-441-6	Sequence 6, Appl	935	5	4.7	355	2	US-08-666-167B-6	Sequence 6, Appl
863	5	4.7	321	2	US-08-478-142-6	Sequence 6, Appl	936	5	4.7	355	3	US-08-890-719-11	Sequence 11, Appl
864	5	4.7	321	2	US-08-478-572-6	Sequence 6, Appl	937	5	4.7	355	3	US-08-890-719-13	Sequence 13, Appl
865	5	4.7	321	4	US-08-484-286-6	Sequence 6, Appl	938	5	4.7	355	4	US-09-143-438-6	Sequence 6, Appl
866	5	4.7	321	5	PCT-US93-08518-6	Sequence 6, Appl	939	5	4.7	355	5	US-09-403-768-6	Sequence 6, Appl
867	5	4.7	322	1	US-08-036-210-11	Sequence 11, Appl	940	5	4.7	355	5	PCT-US93-08528-11	Sequence 11, Appl
868	5	4.7	322	2	US-08-449-609-11	Sequence 11, Appl	941	5	4.7	357	4	US-08-867-611-40	Sequence 40, Appl
869	5	4.7	323	4	US-09-457-646-30	Sequence 30, Appl	942	5	4.7	357	4	US-09-266-464-2	Sequence 2, Appl
870	5	4.7	324	4	US-09-223-938A-31	Sequence 31, Appl	943	5	4.7	358	1	US-08-604-913B-11	Sequence 11, Appl
871	5	4.7	325	1	US-08-087-797-2	Sequence 2, Appl	944	5	4.7	358	1	US-08-433-783-43	Sequence 43, Appl
872	5	4.7	325	1	US-08-142-473A-9	Sequence 9, Appl	945	5	4.7	358	2	US-08-337-358-43	Sequence 43, Appl
873	5	4.7	325	1	US-08-469-203A-9	Sequence 9, Appl	946	5	4.7	358	5	PCT-US95-07537A-43	Sequence 43, Appl
874	5	4.7	325	1	US-08-469-203A-9	Sequence 9, Appl	947	5	4.7	358	5	PCT-US95-07537A-43	Sequence 43, Appl
875	5	4.7	325	3	US-08-984-618-6	Sequence 6, Appl	948	5	4.7	359	2	US-08-560-398-10	Sequence 10, Appl
876	5	4.7	325	3	US-09-154-874-9	Sequence 9, Appl	949	5	4.7	359	4	US-09-134-001C-4842	Sequence 442, Ap
877	5	4.7	327	3	US-09-154-874-8	Sequence 8, Appl	950	5	4.7	359	4	US-09-540-824-2	Sequence 2, Appl
878	5	4.7	327	4	US-09-134-001C-5542	Sequence 5542, Ap	951	5	4.7	362	3	US-09-120-365-74	Sequence 74, Appl
879	5	4.7	329	2	US-08-562-535C-4	Sequence 4, Appl	952	5	4.7	362	4	US-08-867-611-41	Sequence 41, Appl
880	5	4.7	329	4	US-08-742-605D-4	Sequence 4, Appl	953	5	4.7	362	4	US-09-515-039-74	Sequence 74, Appl
881	5	4.7	329	4	US-09-259-294-4	Sequence 4, Appl	954	5	4.7	362	4	US-09-485-648-6	Sequence 6, Appl
882	5	4.7	329	4	US-09-651-200-19	Sequence 19, Appl	955	5	4.7	362	4	US-09-503-565-6	Sequence 6, Appl
883	5	4.7	330	1	US-08-118-270-20	Sequence 20, Appl	956	5	4.7	363	1	US-09-485-649-6	Sequence 6, Appl
884	5	4.7	330	5	US-08-525-864A-4	Sequence 4, Appl	957	5	4.7	363	1	US-08-148-009A-2	Sequence 2, Appl
885	5	4.7	330	5	PCT-US93-08528-20	Sequence 20, Appl	958	5	4.7	363	1	US-08-148-009A-3	Sequence 3, Appl
886	5	4.7	333	4	US-09-230-637-30	Sequence 30, Appl	959	5	4.7	363	2	US-08-697-611-17	Sequence 17, Appl
887	5	4.7	334	4	US-09-457-568-16	Sequence 16, Appl	960	5	4.7	363	2	US-08-697-611-18	Sequence 18, Appl
888	5	4.7	334	4	US-09-457-568-16	Sequence 16, Appl	961	5	4.7	363	2	US-08-697-611-19	Sequence 19, Appl
889	5	4.7	335	4	US-08-961-083-110	Sequence 110, Appl	962	5	4.7	363	2	US-08-697-611-20	Sequence 20, Appl
890	5	4.7	335	4	US-08-983-035A-44	Sequence 44, Appl	963	5	4.7	363	2	US-08-697-611-21	Sequence 21, Appl
891	5	4.7	336	3	US-09-095-163-2	Sequence 2, Appl	964	5	4.7	363	2	US-08-697-611-22	Sequence 22, Appl
892	5	4.7	336	4	US-09-334-601-13	Sequence 13, Appl	965	5	4.7	363	2	US-08-697-611-23	Sequence 23, Appl
893	5	4.7	336	4	US-09-134-001C-5549	Sequence 5549, Ap	966	5	4.7	363	2	US-08-697-611-24	Sequence 24, Appl
894	5	4.7	337	4	US-09-134-001C-3687	Sequence 3687, Ap	967	5	4.7	363	4	US-08-867-611-43	Sequence 43, Appl
895	5	4.7	337	4	US-09-457-646-32	Sequence 32, Appl	968	5	4.7	364	2	US-08-983-035A-26	Sequence 26, Appl
896	5	4.7	338	3	US-08-447-430A-43	Sequence 43, Appl	969	5	4.7	364	2	US-08-204-288-5	Sequence 5, Appl
897	5	4.7	338	3	US-08-890-719-12	Sequence 12, Appl	970	5	4.7	364	4	US-08-978-741-21	Sequence 21, Appl
898	5	4.7	339	3	US-08-707-399B-2	Sequence 2, Appl	971	5	4.7	365	3	US-08-697-611-44	Sequence 44, Appl
899	5	4.7	340	2	US-08-103-385G-18	Sequence 18, Appl	972	5	4.7	365	4	US-08-867-611-42	Sequence 42, Appl
900	5	4.7	340	4	US-09-733-455-6	Sequence 6, Appl	973	5	4.7	365	4	US-09-333-729A-3	Sequence 3, Appl
901	5	4.7	340	4	US-09-733-455-6	Sequence 6, Appl	974	5	4.7	365	4	US-09-457-568-6	Sequence 6, Appl
902	5	4.7	341	1	US-08-356-180-4	Sequence 4, Appl	975	5	4.7	365	5	US-09-457-568-10	Sequence 10, Appl
903	5	4.7	341	4	US-09-134-001C-3497	Sequence 3497, Ap	976	5	4.7	365	4	US-09-457-568-6	Sequence 6, Appl

977 5 4.7 365 4 US-09-457-646-10 Sequence 10, Appl  
978 5 4.7 369 1 US-07-816-283-6 Sequence 6, Appl  
979 5 4.7 369 1 US-07-816-283-6 Sequence 8, Appl  
980 5 4.7 369 1 US-08-417-103-6 Sequence 6, Appl  
981 5 4.7 369 1 US-08-417-103-8 Sequence 8, Appl  
982 5 4.7 369 1 US-08-417-103-16 Sequence 16, Appl  
983 5 4.7 369 2 US-08-411-859-3 Sequence 3, Appl  
984 5 4.7 369 4 US-08-120-601B-9 Sequence 9, Appl  
985 5 4.7 369 4 US-08-387-707-9 Sequence 9, Appl  
986 5 4.7 369 4 US-08-405-271A-9 Sequence 9, Appl  
987 5 4.7 370 1 US-08-118-270-18 Sequence 18, Appl  
988 5 4.7 370 5 PCT-US93-08528-18 Sequence 18, Appl  
989 5 4.7 372 4 US-08-987-943-3 Sequence 3, Appl  
990 5 4.7 372 4 US-07-757-022B-64 Sequence 64, Appl  
991 5 4.7 373 3 US-08-685-466C-2 Sequence 2, Appl  
992 5 4.7 373 4 US-09-586-875-1 Sequence 1, Appl  
993 5 4.7 374 4 US-08-983-035A-28 Sequence 28, Appl  
994 5 4.7 374 4 US-09-134-001C-5551 Sequence 5551, Ap  
995 5 4.7 375 4 US-08-867-611-50 Sequence 50, Appl  
996 5 4.7 376 1 US-08-500-222-6 Sequence 6, Appl  
997 5 4.7 376 2 US-08-500-125-6 Sequence 6, Appl  
998 5 4.7 376 2 US-08-666-367B-8 Sequence 8, Appl  
999 5 4.7 376 2 US-07-778-704B-6 Sequence 6, Appl  
1000 5 4.7 376 4 US-09-143-438-8 Sequence 8, Appl

## ALIGNMENTS

RESULT 1  
US-09-234-332-12  
Sequence 12, Application US/09234332A

Patent No. 6087168  
GENERAL INFORMATION:  
APPLICANT: Cedars-Sinai Medical Center  
APPLICANT: Michel F. Levesque, M.D.  
TITLE OF INVENTION: CONVERSION OF NON-NEURONAL CELLS INTO  
TITLE OF INVENTION: NEURONS; TRANSDIFFERENTIATION OF EPIDERMAL CELLS  
FILE REFERENCE: P07 41494  
CURRENT APPLICATION NUMBER: US/09/234,332A  
CURRENT FILING DATE: 1999-01-20  
NUMBER OF SEQ ID NOS: 16  
SOFTWARE: FastSeq for Windows Version 3.0  
SEQ ID NO 12  
LENGTH: 725  
TYPE: PRT  
ORGANISM: Human  
FEATURE:  
NAME/KEY: PEPTIDE  
LOCATION: (0)...(0)  
OTHER INFORMATION: Wyein transcription factor 1 (My T1); Genbank  
US-09-234-332-12

Query Match 6.6%; Score 7; DB 3; Length 725;  
Best Local Similarity 100.0%; Pred. No. 38;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 19 SGVAVAP 25  
Db 478 SGVAVAP 484

RESULT 2  
US-08-817-787-22  
Sequence 22, Application US/08817787  
Patent No. 6294353  
GENERAL INFORMATION:  
APPLICANT: Pack, Peter  
APPLICANT: Lupas, Andrei  
TITLE OF INVENTION: TARGETED HETERO-ASSOCIATION OF  
TITLE OF INVENTION: RECOMBINANT PROTEINS TO MULTI-FUNCTIONAL COMPLEXES

NUMBER OF SEQUENCES: 36  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: FISH & NEAVE  
STREET: 1251 Avenue of the Americas  
CITY: New York  
STATE: New York  
COUNTRY: USA  
ZIP: 10020  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/817,787  
FILING DATE: 23-SEP-1997  
CLASSIFICATION: 424  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: PCT/EP95/04117  
FILING DATE:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: EP 94 11 6558.1  
FILING DATE: 20-OCT-1994  
ATTORNEY/AGENT INFORMATION:  
NAME: Haley Jr., James F.  
REGISTRATION NUMBER: 27,794  
REFERENCE/DOCKET NUMBER: MORPHO/1  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 212-596-9000  
TELEFAX: 212-596-9090  
INFORMATION FOR SEQ ID NO: 22:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 60 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-817-787-22

Query Match 5.7%; Score 6; DB 4; Length 60;  
Best Local Similarity 100.0%; Pred. No. 44;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 101 GGLKK 106  
Db 28 GGLKK 33

RESULT 3  
US-09-134-001C-4293  
Sequence 4293, Application US/09134001C  
Patent No. 6380370  
GENERAL INFORMATION:  
APPLICANT: Lynn Doucette-Stamm et al  
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS  
TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS  
FILE REFERENCE: GTC-007  
CURRENT APPLICATION NUMBER: US/09/134,001C  
CURRENT FILING DATE: 1998-08-13  
PRIOR APPLICATION NUMBER: US 60/064,964  
PRIOR FILING DATE: 1997-11-08  
PRIOR APPLICATION NUMBER: US 60/055,779  
PRIOR FILING DATE: 1997-08-14  
NUMBER OF SEQ ID NOS: 5674  
SEQ ID NO 4293  
LENGTH: 74  
TYPE: PRT  
ORGANISM: Staphylococcus epidermidis  
US-09-134-001C-4293

Query Match 5.7%; Score 6; DB 4; Length 74;  
Best Local Similarity 100.0%; Pred. No. 53;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;



OY 19 SGVKA 24  
|||||  
DB 4 SGVKA 9

## RESULT 4

US-08-225-757B-14  
Sequence 14, Application US/08225757B.  
Patent No. 5506133  
GENERAL INFORMATION:  
APPLICANT: YU, ET AL.  
TITLE OF INVENTION: Superoxide dismutase-4  
NUMBER OF SEQUENCES: 16  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: CARELLA, BYRNE, BAIN, GILFILLAN,  
STREET: 6 BECKER FARM ROAD  
CITY: ROSELAND  
STATE: NEW JERSEY  
COUNTRY: USA  
ZIP: 07068  
COMPUTER READABLE FORM:  
MEDIUM TYPE: 3.5 INCH DISKETTE  
COMPUTER: IBM PS/2  
OPERATING SYSTEM: MS-DOS  
SOFTWARE: WORD PERFECT 5.1  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/225,757B  
FILING DATE: 11 APR-94  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER:  
ATTORNEY/AGENT INFORMATION:  
FILING DATE:  
NAME: FERRARO, GREGORY D.  
REGISTRATION NUMBER: 36,134  
REFERENCE/DOCKET NUMBER: 325800-106  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 201-994-1700  
TELEFAX: 201-994-1744  
INFORMATION FOR SEQ ID NO: 14:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 80 AMINO ACIDS  
TYPE: AMINO ACID  
STRANDEDNESS:  
TOPOLOGY: LINEAR  
MOLECULE TYPE: PROTEIN  
US-08-225-757B-14

Query Match 5.7%; Score 6; DB 1; Length 80;  
Best Local Similarity 100.0%; Pred. No. 57;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 64 IRVSL 69  
|||||  
DB 48 IRVSL 53

## RESULT 5

US-08-011-398B-16  
Sequence 16, Application US/08011398B  
Patent No. 5512473  
GENERAL INFORMATION:  
APPLICANT: Roger Brent  
APPLICANT: Antonis S. Zervos  
TITLE OF INVENTION: MAX-INTERACTING PROTEINS AND RELATED  
NUMBER OF SEQUENCES: 20  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Fish & Richardson  
STREET: 225 Franklin Street  
CITY: Boston  
STATE: Massachusetts

COUNTRY: U.S.A.  
ZIP: 02110-2804  
COMPUTER READABLE FORM:  
MEDIUM TYPE: 3.5" Diskette, 1.44 MB  
COMPUTER: IBM PS/2 Model 502 or 55SX  
OPERATING SYSTEM: MS-DOS (Version 5.0)  
SOFTWARE: WordPerfect (Version 5.1)  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/011,398B  
FILING DATE: 29 JAN 1993  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER:  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: Paul T. Clark  
REGISTRATION NUMBER: 30,162  
REFERENCE/DOCKET NUMBER: 00786/160001  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (617) 542-5070  
TELEFAX: (617) 542-8906  
TELEX: 200154  
INFORMATION FOR SEQ ID NO: 16:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 80  
TYPE: amino acid  
STRANDEDNESS: N/A  
TOPOLOGY: linear  
US-08-011-398B-16

Query Match 5.7%; Score 6; DB 1; Length 80;  
Best Local Similarity 100.0%; Pred. No. 57;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2 SFLTR 7  
|||||  
DB 19 SFLTR 24

## RESULT 6

US-08-464-051-16  
Sequence 16, Application US/08464051  
Patent No. 5780262  
GENERAL INFORMATION:  
APPLICANT: Roger Brent  
APPLICANT: Antonis S. Zervos  
TITLE OF INVENTION: MAX-INTERACTING PROTEINS AND RELATED  
NUMBER OF SEQUENCES: 20  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Fish & Richardson  
STREET: 225 Franklin Street  
CITY: Boston  
STATE: Massachusetts  
COUNTRY: U.S.A.  
ZIP: 02110-2804  
COMPUTER READABLE FORM:  
MEDIUM TYPE: 3.5" Diskette, 1.44 MB  
COMPUTER: IBM PS/2 Model 502 or 55SX  
OPERATING SYSTEM: MS-DOS (Version 5.0)  
SOFTWARE: WordPerfect (Version 5.1)  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/464,051  
FILING DATE: 05 JUN 1995  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/011,398  
FILING DATE: 29 JAN 1993  
ATTORNEY/AGENT INFORMATION:  
NAME: Paul T. Clark  
REGISTRATION NUMBER: 30,162  
REFERENCE/DOCKET NUMBER: 00786/160002  
TELECOMMUNICATION INFORMATION:

TELEPHONE: (617) 542-5070  
TELEFAX: (617) 542-8906  
TELEX: 200154  
INFORMATION FOR SEQ ID NO: 16:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 80  
TYPE: amino acid  
STRANDEDNESS: N/A  
TOPOLOGY: linear  
US-08-462-051-16

Query Match 5.7%; Score 6; DB 1; Length 80;  
Best Local Similarity 100.0%; Pred. No. 57;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 SFLTLR 7  
DB 19 SFLTLR 24

RESULT 7  
US-08-462-498-16  
Sequence 16, Application US/08462498  
Patent No. 5852169  
GENERAL INFORMATION:  
APPLICANT: Roger Brent  
APPLICANT: Antonis S. Zervos  
TITLE OF INVENTION: MAX-INTERACTING PROTEINS AND RELATED  
NUMBER OF SEQUENCES: 20  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Fish & Richardson  
STREET: 225 Franklin Street  
CITY: Boston  
STATE: Massachusetts  
COUNTRY: U.S.A.  
ZIP: 02110-2804  
COMPUTER READABLE FORM:  
MEDIUM TYPE: 3.5" Diskette, 1.44 Mb  
COMPUTER: IBM PS/2 Model 502 or 55SX  
OPERATING SYSTEM: MS-DOS (Version 5.0)  
SOFTWARE: WordPerfect (Version 5.1)  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/462,498  
FILING DATE:  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/011,398  
FILING DATE: 29 JAN 1993  
ATTORNEY/AGENT INFORMATION:  
NAME: Paul T. Clark  
REGISTRATION NUMBER: 30,162  
REFERENCE/DOCKET NUMBER: 00786/160001  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (617) 542-5070  
TELEFAX: (617) 542-8906  
TELEX: 200154  
INFORMATION FOR SEQ ID NO: 16:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 80  
TYPE: amino acid  
STRANDEDNESS: N/A  
TOPOLOGY: linear  
US-08-462-498-16

Query Match 5.7%; Score 6; DB 2; Length 80;  
Best Local Similarity 100.0%; Pred. No. 57;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 SFLTLR 7  
DB 19 SFLTLR 24

RESULT 8  
US-08-554-385-15  
Sequence 15, Application US/08554385  
Patent No. 6017692  
GENERAL INFORMATION:  
APPLICANT: Roger Brent  
APPLICANT: Antonis S. Zervos  
TITLE OF INVENTION: MAX-INTERACTING PROTEINS AND RELATED  
NUMBER OF SEQUENCES: 32  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Fish & Richardson P.C.  
STREET: 225 Franklin Street  
CITY: Boston  
STATE: Massachusetts  
COUNTRY: U.S.A.  
ZIP: 02110-2804  
COMPUTER READABLE FORM:  
MEDIUM TYPE: 3.5" Diskette, 1.44 Mb  
COMPUTER: IBM PS/2 Model 502 or 55SX  
OPERATING SYSTEM: MS-DOS (Version 5.0)  
SOFTWARE: WordPerfect (Version 5.1)  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/554,385  
FILING DATE: No. 6017692ember 8, 1995  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER:  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: Karen F. Lech  
REGISTRATION NUMBER: 35,238  
REFERENCE/DOCKET NUMBER: 00786/252001  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (617) 542-5070  
TELEFAX: (617) 542-8906  
TELEX: 200154  
INFORMATION FOR SEQ ID NO: 15:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 80 amino acids  
TYPE: amino acid  
STRANDEDNESS: not relevant  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-554-385-15

Query Match 5.7%; Score 6; DB 3; Length 80;  
Best Local Similarity 100.0%; Pred. No. 57;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 SFLTLR 7  
DB 19 SFLTLR 24

RESULT 9  
US-09-142-078-66  
Sequence 66, Application US/09142078  
Patent No. 6172041  
GENERAL INFORMATION:  
APPLICANT: McCabe, R. Tyler  
APPLICANT: Zhou, Li-Ming  
APPLICANT: Layer, Richard T.  
TITLE OF INVENTION: Use of Conantokins  
NUMBER OF SEQUENCES: 71  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Rothwell, Figg, Ernst & Kurtz, P.C.  
STREET: 555 Thirteenth Street, N.W., Suite 701-E  
CITY: Washington  
STATE: D.C.  
COUNTRY: USA  
ZIP: 20004

```
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/142,078
FILING DATE: 10-FEB-1999
PRIOR APPLICATION DATA:
APPLICATION NUMBER: WO US97/12652
FILING DATE: 21-JUL-1997
APPLICATION NUMBER: US 08/762,377
FILING DATE: 06-DEC-1996
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/684,750
FILING DATE: 22-JUL-1996
ATTORNEY/AGENT INFORMATION:
NAME: Ihnen, Jeffrey L.
REGISTRATION NUMBER: 28,957
REFERENCE/DOCKET NUMBER: 2314-135.A
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202-783-6040
TELEFAX: 202-783-6031
INFORMATION FOR SEQ ID NO: 66:
SEQUENCE CHARACTERISTICS:
LENGTH: 95 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-09-142-078-66

Query Match      5.7%; Score 6; DB 4; Length 95;
Best Local Similarity 100.0%; Pred. No. 66;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      90 DDNDRD 95
      |||||
Db      51 DDNDRD 56

RESULT 10
US-09-357-141-66
Sequence 66, Application US/09357141
Patent No. 6277825
GENERAL INFORMATION:
APPLICANT: Oliveira, Baldomero M.
APPLICANT: McIntosh, J. Michael
APPLICANT: McCabe, R. Tyler
APPLICANT: Zhou, Richard T.
TITLE OF INVENTION: Use of Conantokins for Treating Pain
FILE REFERENCE: 2314-171
CURRENT APPLICATION NUMBER: US/09/357,141
CURRENT FILING DATE: 1999-07-20
PRIOR APPLICATION NUMBER: US 09/283,277
PRIOR FILING DATE: 1999-04-01
PRIOR APPLICATION NUMBER: US 09/142,078
PRIOR FILING DATE: 1999-02-10
PRIOR APPLICATION NUMBER: WO US97/12652
PRIOR FILING DATE: 1997-07-21
PRIOR APPLICATION NUMBER: US 08/762,377
PRIOR FILING DATE: 1996-12-06
PRIOR APPLICATION NUMBER: US 08/684,750
PRIOR FILING DATE: 1996-07-22
NUMBER OF SEQ ID NOS: 71
SOFTWARE: Patentin Ver. 2.0
SEQ ID NO 66
LENGTH: 95
TYPE: PRT
ORGANISM: Conus quercinus
US-09-357-141-66
```

```
Query Match      5.7%; Score 6; DB 4; Length 95;
Best Local Similarity 100.0%; Pred. No. 66;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      90 DDNDRD 95
      |||||
Db      51 DDNDRD 56

RESULT 11
US-09-533-889-66
Sequence 66, Application US/09533889
Patent No. 639574
GENERAL INFORMATION:
APPLICANT: McCabe, R. Tyler
APPLICANT: Zhou, Li-Ming
APPLICANT: Layer, Richard T.
APPLICANT: Oliveira, Baldomero M.
APPLICANT: McIntosh, J. Michael
TITLE OF INVENTION: Use of Conantokins
NUMBER OF SEQUENCES: 71
CORRESPONDENCE ADDRESS:
ADDRESSEE: Rothwell, Figg, Ernst & Kurz, P.C.
STREET: 555 Thirteenth Street, N.W., Suite 701-E
CITY: Washington
STATE: D.C.
COUNTRY: USA
ZIP: 20004
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/533,889
FILING DATE: 22 MAR-2000
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 09/142,078
FILING DATE: 10-FEB-1999
PRIOR APPLICATION DATA:
APPLICATION NUMBER: WO US97/12652
FILING DATE: 21-JUL-1997
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/762,377
FILING DATE: 06-DEC-1996
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/684,750
FILING DATE: 22-JUL-1996
ATTORNEY/AGENT INFORMATION:
NAME: Ihnen, Jeffrey L.
REGISTRATION NUMBER: 28,957
REFERENCE/DOCKET NUMBER: 2314-168.A
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202-783-6040
TELEFAX: 202-783-6031
INFORMATION FOR SEQ ID NO: 66:
SEQUENCE CHARACTERISTICS:
LENGTH: 95 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-09-533-889-66

Query Match      5.7%; Score 6; DB 4; Length 95;
Best Local Similarity 100.0%; Pred. No. 66;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      90 DDNDRD 95
      |||||
Db      51 DDNDRD 56

RESULT 12
```

US-08-710-330A-1  
; Sequence 1, Application US/08710330A  
; Patent No. 5854041  
; GENERAL INFORMATION:  
; APPLICANT: Brayer, Gary D.  
; APPLICANT: Lee, Hung  
; APPLICANT: Mauk, Grant A.  
; APPLICANT: Smith, Michael  
; APPLICANT: Tong, Harry  
; APPLICANT: Wan, Lianglu  
; TITLE OF INVENTION: MYOGLOBIN WITH PEROXIDASE ACTIVITY  
; NUMBER OF SEQUENCES: 11  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Campbell & Flores LLP  
; STREET: 4370 La Jolla Village Drive, Suite 700  
; CITY: San Diego  
; STATE: California  
; COUNTRY: USA  
; ZIP: 92122  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentin Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/710.330A  
; FILING DATE: 16-SEP-1996  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: GAY, David A.  
; REGISTRATION NUMBER: 39,200  
; REFERENCE/DOCKET NUMBER: P-SM 2262  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (619) 535-9001  
; TELEFAX: (619) 535-8949  
; INFORMATION FOR SEQ ID NO: 1:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 113 amino acids  
; TYPE: amino acid  
; STRANDEDNESS:  
; TOPOLOGY: linear  
US-08-710-330A-1  
; Query Match. 5.7%; Score 6; DB 2; Length 113;  
; Best Local Similarity 100.0%; Pred. No. 77;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 101 GGLKK 106  
Db 73 GGLKK 78  
RESULT 13  
US-08-710-330A-3  
; Sequence 3, Application US/08710330A  
; Patent No. 5854041  
; GENERAL INFORMATION:  
; APPLICANT: Brayer, Gary D.  
; APPLICANT: Lee, Hung  
; APPLICANT: Mauk, Grant A.  
; APPLICANT: Smith, Michael  
; APPLICANT: Tong, Harry  
; APPLICANT: Wan, Lianglu  
; TITLE OF INVENTION: MYOGLOBIN WITH PEROXIDASE ACTIVITY  
; NUMBER OF SEQUENCES: 11  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Campbell & Flores LLP  
; STREET: 4370 La Jolla Village Drive, Suite 700  
; CITY: San Diego  
; STATE: California  
; COUNTRY: USA  
; ZIP: 92122  
; COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentin Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/710.330A  
; FILING DATE: 16-SEP-1996  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: GAY, David A.  
; REGISTRATION NUMBER: 39,200  
; REFERENCE/DOCKET NUMBER: P-SM 2262  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (619) 535-9001  
; TELEFAX: (619) 535-8949  
; INFORMATION FOR SEQ ID NO: 3:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 113 amino acids  
; TYPE: amino acid  
; STRANDEDNESS:  
; TOPOLOGY: linear  
US-08-710-330A-3  
; Query Match. 5.7%; Score 6; DB 2; Length 113;  
; Best Local Similarity 100.0%; Pred. No. 77;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 101 GGLKK 106  
Db 73 GGLKK 78  
RESULT 14  
US-08-710-330A-4  
; Sequence 4, Application US/08710330A  
; Patent No. 5854041  
; GENERAL INFORMATION:  
; APPLICANT: Brayer, Gary D.  
; APPLICANT: Lee, Hung  
; APPLICANT: Mauk, Grant A.  
; APPLICANT: Smith, Michael  
; APPLICANT: Tong, Harry  
; APPLICANT: Wan, Lianglu  
; TITLE OF INVENTION: MYOGLOBIN WITH PEROXIDASE ACTIVITY  
; NUMBER OF SEQUENCES: 11  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Campbell & Flores LLP  
; STREET: 4370 La Jolla Village Drive, Suite 700  
; CITY: San Diego  
; STATE: California  
; COUNTRY: USA  
; ZIP: 92122  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentin Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/710.330A  
; FILING DATE: 16-SEP-1996  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: GAY, David A.  
; REGISTRATION NUMBER: 39,200  
; REFERENCE/DOCKET NUMBER: P-SM 2262  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (619) 535-9001  
; TELEFAX: (619) 535-8949  
; INFORMATION FOR SEQ ID NO: 4:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 113 amino acids  
; TYPE: amino acid  
; STRANDEDNESS:

```
; TOPOLOGY: linear
US-08-710-330A-4
Query Match          5.7%; Score 6; DB 2; Length 113;
Best Local Similarity 100.0%; Pred.No.77;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 101 GILKK 106
   |||||
Db 73 GILKK 78

RESULT 15
US-08-710-330A-5
; Sequence 5, Application US/08710330A
; Patent No. 5854041
; GENERAL INFORMATION:
; APPLICANT: Brayer, Gary D.
; APPLICANT: Lee, Hung
; APPLICANT: Mauk, Grant A.
; APPLICANT: Smith, Michael
; APPLICANT: Tong, Harry
; APPLICANT: Wan, Lianglu
; TITLE OF INVENTION: MYOGLOBIN WITH PEROXIDASE ACTIVITY
; NUMBER OF SEQUENCES: 11
; CORRESPONDENCE ADDRESS:
; ADDRESS: Campbell & Flores LLP
; STREET: 4370 La Jolla Village Drive, Suite 700
; CITY: San Diego
; STATE: California
; COUNTRY: USA
; ZIP: 92122
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/710,330A
; FILING DATE: 16-SEP-1996
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: GAY, David A.
; REGISTRATION NUMBER: 39,200
; REFERENCE/DOCKET NUMBER: P-SM 2262
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (619) 535-9001
; TELEFAX: (619) 535-8949
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 113 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
US-08-710-330A-5
Query Match          5.7%; Score 6; DB 2; Length 113;
Best Local Similarity 100.0%; Pred.No.77;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 101 GILKK 106
   |||||
Db 73 GILKK 78

RESULT 16
US-08-710-330A-11
; Sequence 11, Application US/08710330A
; Patent No. 5854041
; GENERAL INFORMATION:
; APPLICANT: Brayer, Gary D.
; APPLICANT: Lee, Hung
; APPLICANT: Mauk, Grant A.

; APPLICANT: Smith, Michael
; APPLICANT: Tong, Harry
; APPLICANT: Wan, Lianglu
; TITLE OF INVENTION: MYOGLOBIN WITH PEROXIDASE ACTIVITY
; NUMBER OF SEQUENCES: 11
; CORRESPONDENCE ADDRESS:
; ADDRESS: Campbell & Flores LLP
; STREET: 4370 La Jolla Village Drive, Suite 700
; CITY: San Diego
; STATE: California
; COUNTRY: USA
; ZIP: 92122
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/710,330A
; FILING DATE: 16-SEP-1996
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: GAY, David A.
; REGISTRATION NUMBER: 39,200
; REFERENCE/DOCKET NUMBER: P-SM 2262
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (619) 535-9001
; TELEFAX: (619) 535-8949
; INFORMATION FOR SEQ ID NO: 11:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 154 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
US-08-710-330A-11
Query Match          5.7%; Score 6; DB 2; Length 154;
Best Local Similarity 100.0%; Pred.No.1e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 101 GILKK 106
   |||||
Db 74 GILKK 79

RESULT 17
US-08-817-787-18
; Sequence 18, Application US/08817787
; Patent No. 6294353
; GENERAL INFORMATION:
; APPLICANT: Pack, Peter
; APPLICANT: Lupas, Andrei
; TITLE OF INVENTION: TARGETED HETERO-ASSOCIATION OF
; RECOMBINANT PROTEINS TO MULTI-FUNCTIONAL COMPLEXES
; NUMBER OF SEQUENCES: 36
; CORRESPONDENCE ADDRESS:
; ADDRESS: FISH & NEAVE
; STREET: 1251 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10020
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/817,787
; FILING DATE: 23-SEP-1997
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/EP95/04117
```

;; FILING DATE:  
;; PRIOR APPLICATION DATA:  
;; APPLICATION NUMBER: EP 94 11 6558.1  
;; FILING DATE: 20-OCT-1994  
;; ATTORNEY/AGENT INFORMATION:  
;; NAME: Haley Jr., James F.  
;; REGISTRATION NUMBER: 27,794  
;; REFERENCE/DOCKET NUMBER: MORPHO/1  
;; TELECOMMUNICATION INFORMATION:  
;; TELEPHONE: 212-596-9000  
;; TELEFAX: 212-596-9090  
;; INFORMATION FOR SEQ ID NO: 18:  
;; SEQUENCE CHARACTERISTICS:  
;; LENGTH: 154 amino acids  
;; TYPE: amino acid  
;; STRANDEDNESS:  
;; TOPOLOGY: linear  
;; MOLECULE TYPE: protein  
;; ORIGINAL SOURCE:  
;; ORGANISM: Homo sapiens  
US-08-817-787-18

Query Match 5.7%; Score 6; DB 4; Length 154;  
Best Local Similarity 100.0%; Pred. No. 1e+02;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 101 GGILKK 106  
|||  
Db 74 GGILKK 79

RESULT 18  
US-09-512-342-20  
; Sequence 20, Application US/09512342  
; Patent No. 6388068  
; GENERAL INFORMATION:  
; APPLICANT: SATOH, SHINOBU  
; APPLICANT: MASUDA, SUSUMU  
; TITLE OF INVENTION: METHOD FOR PRODUCING FOREIGN POLYPEPTIDE IN PLANT  
; TITLE OF INVENTION: INTERCELLULAR FLUID  
; FILE REFERENCE: 081356/0142  
; CURRENT APPLICATION NUMBER: US/09/512,342  
; CURRENT FILING DATE: 2000-02-24  
; NUMBER OF SEQ ID NOS: 38  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 20  
; LENGTH: 168  
; TYPE: PRT  
; ORGANISM: Cucumis sativus  
US-09-512-342-20

Query Match 5.7%; Score 6; DB 4; Length 168;  
Best Local Similarity 100.0%; Pred. No. 1.1e+02;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 85 VGYGRD 90  
|||  
Db 53 VGYGRD 58

RESULT 19  
US-09-199-637A-158  
; Sequence 158, Application US/09199637A  
; Patent No. 6355411  
; GENERAL INFORMATION:  
; APPLICANT: Ausubel, Frederick  
; APPLICANT: Goodman, Howard W.  
; APPLICANT: Rahme, Laurence G.  
; APPLICANT: Mahajan-Miklos, Shalina  
; APPLICANT: Tan, Man-Wah  
; APPLICANT: Cao, Hui  
; APPLICANT: Drenthard, Eliana  
; APPLICANT: Tsongalis, John

;; TITLE OF INVENTION: VIRULENCE-ASSOCIATED NUCLEIC ACID  
;; TITLE OF INVENTION: SEQUENCES AND USES THEREOF  
;; FILE REFERENCE: 00786/361002  
;; CURRENT APPLICATION NUMBER: US/09/199,637A  
;; CURRENT FILING DATE: 1998-11-25  
;; PRIOR APPLICATION NUMBER: 60/066,517  
;; PRIOR FILING DATE: 1997-11-25  
;; NUMBER OF SEQ ID NOS: 437  
;; SOFTWARE: FastSeq for Windows Version 4.0  
;; SEQ ID NO 158  
;; LENGTH: 170  
;; TYPE: PRT  
;; ORGANISM: Pseudomonas aeruginosa  
US-09-199-637A-158

Query Match 5.7%; Score 6; DB 4; Length 170;  
Best Local Similarity 100.0%; Pred. No. 1.1e+02;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 34 ARORRP 39  
|||  
Db 65 ARORRP 70

RESULT 20  
US-08-531-525-45  
; Sequence 45, Application US/08531525  
; Patent No. 5840683  
; GENERAL INFORMATION:  
; APPLICANT: Hlavka, Joseph J.  
; APPLICANT: Pincus, Matthew R.  
; APPLICANT: No. 5840683le, John F.  
; APPLICANT: Abajian, Henry B.  
; APPLICANT: Kende, Andrew S.  
; TITLE OF INVENTION: Peptides Inhibiting the Oncogenic Action  
; TITLE OF INVENTION: OF P21 Ras  
; NUMBER OF SEQUENCES: 52  
; CORRESPONDENCE ADDRESSES:  
; ADDRESSEE: Greenlee and Winner, P.C.  
; STREET: 5370 Manhattan Circle, Suite 201  
; CITY: Boulder  
; STATE: Colorado  
; COUNTRY: US  
; ZIP: 80303  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/531,525  
; FILING DATE: 21-SEP-1995  
; CLASSIFICATION: 530  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Feider, Donna M.  
; REGISTRATION NUMBER: 33,878  
; REFERENCE/DOCKET NUMBER: 37-94  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (303) 499-8080  
; TELEFAX: (303) 499-8089  
; INFORMATION FOR SEQ ID NO: 45:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 190 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; HYPOTHETICAL: NO  
; ORIGINAL SOURCE:  
; ORGANISM: Dictyostelium discoideum  
US-08-531-525-45

Query Match 5.7%; Score 6; DB 2; Length 190;

Best Local Similarity 100.0%; Pred. No. 1.2e+02;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Cy 101 GILKK 106  
Db 176 GILKK 181

RESULT 21  
US-08-718-270A-45  
Sequence 45, Application US/08718270A  
Patent No. 5910478  
GENERAL INFORMATION:  
APPLICANT: Hlavka, Joseph J.  
APPLICANT: Pincus, Matthew R.  
APPLICANT: No. 59104781e, John F.  
APPLICANT: Abajian, Henry B.  
APPLICANT: Kende, Andrew S.  
TITLE OF INVENTION: Peptidomimetics Inhibiting  
TITLE OF INVENTION: the Oncogenic Action of p21 Ras  
NUMBER OF SEQUENCES: 52  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Greenlee, Winner and Sullivan, P.C.  
STREET: 5370 Manhattan Circle, Suite 201  
CITY: Boulder  
STATE: Colorado  
COUNTRY: US  
ZIP: 80303  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/718,270A  
FILING DATE: 20-SEP-1996  
CLASSIFICATION: 514  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/531,525  
FILING DATE: 21-SEP-1995  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 60/004,091  
FILING DATE: 21-SEP-1995  
ATTORNEY/AGENT INFORMATION:  
NAME: Feiber, Donna M.  
REGISTRATION NUMBER: 33,878  
REFERENCE/DOCKET NUMBER: 78-95  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (303) 499-8080  
TELEFAX: (303) 499-8089  
INFORMATION FOR SEQ ID NO: 45:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 190 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: Protein  
HYPOTHETICAL: NO  
ORIGINAL SOURCE:  
ORGANISM: Dictyostelium discoideum  
US-08-718-270A-45

Query Match 5.7%; Score 6; DB 2; Length 190;  
Best Local Similarity 100.0%; Pred. No. 1.2e+02;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Cy 101 GILKK 106  
Db 176 GILKK 181  
RESULT 22  
US-09-170-769A-4

Sequence 4, Application US/09170769A  
Patent No. 6444206  
GENERAL INFORMATION:  
APPLICANT: THE SCRIPPS RESEARCH INSTITUTE  
APPLICANT: LETURCO, Didier  
APPLICANT: MORIARTY, Ann  
APPLICANT: ULEVITCH, Richard  
APPLICANT: TOBIAS, Peter  
TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR INHIBITING CD14 MEDIATED CELL ACTIV  
FILE REFERENCE: SCRIPI40-3  
CURRENT APPLICATION NUMBER: US/09/170,769A  
CURRENT FILING DATE: 1998-10-13  
PRIOR APPLICATION NUMBER: US 08/070,160  
PRIOR FILING DATE: 1993-05-28  
NUMBER OF SEQ ID NOS: 26  
SOFTWARE: PatentIn version 3.0  
SEQ ID NO 4  
LENGTH: 215  
TYPE: PRT  
ORGANISM: Murine  
US-09-170-769A-4

Query Match 5.7%; Score 6; DB 4; Length 215;  
Best Local Similarity 100.0%; Pred. No. 1.3e+02;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Cy 59 NLQSGI 64  
Db 54 NLQSGI 59

RESULT 23  
US-07-706-691G-1  
Sequence 1, Application US/07706691G  
Patent No. 5340735  
GENERAL INFORMATION:  
APPLICANT: Goddette, Dean W  
TITLE OF INVENTION: Third Generation Protease Mutants  
NUMBER OF SEQUENCES: 104  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Henkel Corporation  
STREET: 140 Germantown Pike, Suite 150  
CITY: Plymouth Meeting  
STATE: Pennsylvania  
COUNTRY: USA  
ZIP: 19462  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/07/706,691G  
FILING DATE: 19910529  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Drach, John E.  
REGISTRATION NUMBER: 32891  
REFERENCE/DOCKET NUMBER: M4922  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 215-832-2215  
TELEFAX: 215-941-6067  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 269 amino acids  
TYPE: AMINO ACID  
TOPOLOGY: unknown  
MOLECULE TYPE: protein  
HYPOTHETICAL: NO  
ANTI-SENSE: NO  
ORIGINAL SOURCE:  
ORGANISM: Serpine Protease  
STRAIN: Bacillus lentus DSM 5843

IMMEDIATE SOURCE:  
CLONE: S3T, V41, A188P, V193M, V1991  
US-07-706-691G-1

Query Match 5.7%; Score 6; DB 1; Length 269;  
Best Local Similarity 100.0%; Pred. No. 1.6e+02;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 19 SGVKA 24  
DB 24 SGVKA 29

RESULT 24  
US-07-706-691G-2  
Sequence 2, Application US/07706691G  
Patent No. 5340735  
GENERAL INFORMATION:  
APPLICANT: Goddette, Dean W  
TITLE OF INVENTION: Third Generation Protease Mutants  
NUMBER OF SEQUENCES: 104  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Henkel Corporation  
STREET: 140 Germantown Pike, Suite 150  
CITY: Plymouth Meeting  
STATE: Pennsylvania  
COUNTRY: USA  
ZIP: 19462  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/07/706,691G  
FILING DATE: 19910529  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Drach, John E.  
REGISTRATION NUMBER: 32891  
REFERENCE/DOCKET NUMBER: M4922  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 215-832-2215  
TELEFAX: 215-941-6067  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 269 amino acids  
TYPE: AMINO ACID  
TOPOLOGY: unknown  
MOLECULE TYPE: protein  
HYPOTHETICAL: NO  
ANTI-SENSE: NO  
ORIGINAL SOURCE:  
ORGANISM: Serine Protease  
STRAIN: Bacillus lentus DSM 5843  
IMMEDIATE SOURCE:  
CLONE: S3T, A188P, V193M, V1991  
US-07-706-691G-2

Query Match 5.7%; Score 6; DB 1; Length 269;  
Best Local Similarity 100.0%; Pred. No. 1.6e+02;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 19 SGVKA 24  
DB 24 SGVKA 29

RESULT 25  
US-07-706-691G-3  
Sequence 3, Application US/07706691G  
Patent No. 5340735  
GENERAL INFORMATION:

APPLICANT: Goddette, Dean W  
TITLE OF INVENTION: Third Generation Protease Mutants  
NUMBER OF SEQUENCES: 104  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Henkel Corporation  
STREET: 140 Germantown Pike, Suite 150  
CITY: Plymouth Meeting  
STATE: Pennsylvania  
COUNTRY: USA  
ZIP: 19462  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/07/706,691G  
FILING DATE: 19910529  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Drach, John E.  
REGISTRATION NUMBER: 32891  
REFERENCE/DOCKET NUMBER: M4922  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 215-832-2215  
TELEFAX: 215-941-6067  
INFORMATION FOR SEQ ID NO: 3:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 269 amino acids  
TYPE: AMINO ACID  
TOPOLOGY: unknown  
MOLECULE TYPE: protein  
HYPOTHETICAL: NO  
ANTI-SENSE: NO  
ORIGINAL SOURCE:  
ORGANISM: Serine Protease  
STRAIN: Bacillus lentus DSM 5843  
IMMEDIATE SOURCE:  
CLONE: V41, A188P, V193M, V1991  
US-07-706-691G-3

Query Match 5.7%; Score 6; DB 1; Length 269;  
Best Local Similarity 100.0%; Pred. No. 1.6e+02;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 19 SGVKA 24  
DB 24 SGVKA 29

RESULT 26  
US-07-706-691G-4  
Sequence 4, Application US/07706691G  
Patent No. 5340735  
GENERAL INFORMATION:  
APPLICANT: Goddette, Dean W  
TITLE OF INVENTION: Third Generation Protease Mutants  
NUMBER OF SEQUENCES: 104  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Henkel Corporation  
STREET: 140 Germantown Pike, Suite 150  
CITY: Plymouth Meeting  
STATE: Pennsylvania  
COUNTRY: USA  
ZIP: 19462  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/07/706,691G  
FILING DATE: 19910529



```

; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Drach, John E.
; REGISTRATION NUMBER: 32891
; REFERENCE/DOCKET NUMBER: M4922
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 215-832-2215
; TELEFAX: 215-941-6067
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 269 amino acids
; TYPE: AMINO ACID
; TOPOLOGY: unknown
; MOLECULE TYPE: protein
; HYPOTHEICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: Serine Protease
; STRAIN: Bacillus lentus DSM 5843
; IMMEDIATE SOURCE:
; CLONE: S139Y, A188P, V193M, V199I
US-07-706-691G-4

Query Match          5.7%; Score 6; DB 1; Length 269;
Best Local Similarity 100.0%; Pred. No. 1.6e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 19 SGVKVA 24
Db 24 SGVKVA 29

RESULT 27
US-07-706-691G-5
; Sequence 5, Application US/07706691G
; Patent No. 5340735
; GENERAL INFORMATION:
; APPLICANT: Goddette, Dean W
; TITLE OF INVENTION: Third Generation Protease Mutants
; NUMBER OF SEQUENCES: 104
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Henkel Corporation
; STREET: 140 Germantown Pike, Suite 150
; CITY: Plymouth Meeting
; STATE: Pennsylvania
; COUNTRY: USA
; ZIP: 19462
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/706,691G
; FILING DATE: 19910529
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Drach, John E.
; REGISTRATION NUMBER: 32891
; REFERENCE/DOCKET NUMBER: M4922
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 215-832-2215
; TELEFAX: 215-941-6067
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 269 amino acids
; TYPE: AMINO ACID
; TOPOLOGY: unknown
; MOLECULE TYPE: protein
; HYPOTHEICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: Serine Protease
; STRAIN: Bacillus lentus DSM 5843
; IMMEDIATE SOURCE:
; CLONE: A188P, V193M, V199I
US-07-706-691G-6

Query Match          5.7%; Score 6; DB 1; Length 269;
Best Local Similarity 100.0%; Pred. No. 1.6e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 19 SGVKVA 24
Db 24 SGVKVA 29

RESULT 28
US-07-706-691G-6
; Sequence 6, Application US/07706691G
; Patent No. 5340735
; GENERAL INFORMATION:
; APPLICANT: Goddette, Dean W
; TITLE OF INVENTION: Third Generation Protease Mutants
; NUMBER OF SEQUENCES: 104
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Henkel Corporation
; STREET: 140 Germantown Pike, Suite 150
; CITY: Plymouth Meeting
; STATE: Pennsylvania
; COUNTRY: USA
; ZIP: 19462
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/706,691G
; FILING DATE: 19910529
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Drach, John E.
; REGISTRATION NUMBER: 32891
; REFERENCE/DOCKET NUMBER: M4922
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 215-832-2215
; TELEFAX: 215-941-6067
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 269 amino acids
; TYPE: AMINO ACID
; TOPOLOGY: unknown
; MOLECULE TYPE: protein
; HYPOTHEICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: Serine Protease
; STRAIN: Bacillus lentus DSM 5843
; IMMEDIATE SOURCE:
; CLONE: A188P, V193M, V199I
US-07-706-691G-7

Query Match          5.7%; Score 6; DB 1; Length 269;
Best Local Similarity 100.0%; Pred. No. 1.6e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 19 SGVKVA 24
Db 24 SGVKVA 29

RESULT 29
US-07-706-691G-7
; Sequence 7, Application US/07706691G
; Patent No. 5340735
```

```

; STRAIN: Bacillus lentus DSM 5843
; IMMEDIATE SOURCE:
; CLONE: S130T, S139Y, A188P, V193M, V199I
US-07-706-691G-5

Query Match          5.7%; Score 6; DB 1; Length 269;
Best Local Similarity 100.0%; Pred. No. 1.6e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 19 SGVKVA 24
Db 24 SGVKVA 29

RESULT 26
US-07-706-691G-5
; Sequence 5, Application US/07706691G
; Patent No. 5340735
; GENERAL INFORMATION:
; APPLICANT: Goddette, Dean W
; TITLE OF INVENTION: Third Generation Protease Mutants
; NUMBER OF SEQUENCES: 104
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Henkel Corporation
; STREET: 140 Germantown Pike, Suite 150
; CITY: Plymouth Meeting
; STATE: Pennsylvania
; COUNTRY: USA
; ZIP: 19462
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/706,691G
; FILING DATE: 19910529
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Drach, John E.
; REGISTRATION NUMBER: 32891
; REFERENCE/DOCKET NUMBER: M4922
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 215-832-2215
; TELEFAX: 215-941-6067
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 269 amino acids
; TYPE: AMINO ACID
; TOPOLOGY: unknown
; MOLECULE TYPE: protein
; HYPOTHEICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: Serine Protease
; STRAIN: Bacillus lentus DSM 5843
; IMMEDIATE SOURCE:
; CLONE: A188P, V193M, V199I
US-07-706-691G-6

Query Match          5.7%; Score 6; DB 1; Length 269;
Best Local Similarity 100.0%; Pred. No. 1.6e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 19 SGVKVA 24
Db 24 SGVKVA 29

RESULT 27
US-07-706-691G-6
; Sequence 6, Application US/07706691G
; Patent No. 5340735
; GENERAL INFORMATION:
; APPLICANT: Goddette, Dean W
; TITLE OF INVENTION: Third Generation Protease Mutants
; NUMBER OF SEQUENCES: 104
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Henkel Corporation
; STREET: 140 Germantown Pike, Suite 150
; CITY: Plymouth Meeting
; STATE: Pennsylvania
; COUNTRY: USA
; ZIP: 19462
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/706,691G
; FILING DATE: 19910529
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Drach, John E.
; REGISTRATION NUMBER: 32891
; REFERENCE/DOCKET NUMBER: M4922
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 215-832-2215
; TELEFAX: 215-941-6067
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 269 amino acids
; TYPE: AMINO ACID
; TOPOLOGY: unknown
; MOLECULE TYPE: protein
; HYPOTHEICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: Serine Protease
; STRAIN: Bacillus lentus DSM 5843
; IMMEDIATE SOURCE:
; CLONE: A188P, V193M, V199I
US-07-706-691G-7

Query Match          5.7%; Score 6; DB 1; Length 269;
Best Local Similarity 100.0%; Pred. No. 1.6e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 19 SGVKVA 24
Db 24 SGVKVA 29

RESULT 28
US-07-706-691G-7
; Sequence 7, Application US/07706691G
; Patent No. 5340735
```

```

; GENERAL INFORMATION:
; APPLICANT: Goddette, Dean W
; TITLE OF INVENTION: Third Generation Protease Mutants
; NUMBER OF SEQUENCES: 104
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Henkel Corporation
; STREET: 140 Germantown Pike, Suite 150
; CITY: Plymouth Meeting
; STATE: Pennsylvania
; COUNTRY: USA
; ZIP: 19462
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/706,691G
; FILING DATE: 19910529
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Drach, John E.
; REGISTRATION NUMBER: 32891
; REFERENCE/DOCKET NUMBER: M4922
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 215-941-6067
; TELEFAX: 215-941-6067
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 269 amino acids
; TYPE: AMINO ACID
; TOPOLOGY: unknown
; MOLECULE TYPE: protein
; HYPOTHEICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: Serine Protease
; STRAIN: Bacillus lentus DSM 5843
; IMMEDIATE SOURCE:
; CLONE: S3T, A188P, V193M
;
US-07-706-691G-7
;
Query Match 5.7%; Score 6; DB 1; Length 269;
Best Local Similarity 100.0%; Pred. No. 1.6e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 19 SGVXVA 24
Db 24 SGVXVA 29

RESULT 30
US-07-706-691G-8
; Sequence 8, Application US/07706691G
; Patent No. 5340735
; GENERAL INFORMATION:
; APPLICANT: Goddette, Dean W
; TITLE OF INVENTION: Third Generation Protease Mutants
; NUMBER OF SEQUENCES: 104
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Henkel Corporation
; STREET: 140 Germantown Pike, Suite 150
; CITY: Plymouth Meeting
; STATE: Pennsylvania
; COUNTRY: USA
; ZIP: 19462
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/706,691G
```

```

; FILING DATE: 19910529
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Drach, John E.
; REGISTRATION NUMBER: 32891
; REFERENCE/DOCKET NUMBER: M4922
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 215-941-6067
; TELEFAX: 215-941-6067
; INFORMATION FOR SEQ ID NO: 8:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 269 amino acids
; TYPE: AMINO ACID
; TOPOLOGY: unknown
; MOLECULE TYPE: protein
; HYPOTHEICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: Serine Protease
; STRAIN: Bacillus lentus DSM 5843
; IMMEDIATE SOURCE:
; CLONE: S157T
;
US-07-706-691G-8
;
Query Match 5.7%; Score 6; DB 1; Length 269;
Best Local Similarity 100.0%; Pred. No. 1.6e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 19 SGVXVA 24
Db 24 SGVXVA 29

RESULT 31
US-07-706-691G-9
; Sequence 9, Application US/07706691G
; Patent No. 5340735
; GENERAL INFORMATION:
; APPLICANT: Goddette, Dean W
; TITLE OF INVENTION: Third Generation Protease Mutants
; NUMBER OF SEQUENCES: 104
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Henkel Corporation
; STREET: 140 Germantown Pike, Suite 150
; CITY: Plymouth Meeting
; STATE: Pennsylvania
; COUNTRY: USA
; ZIP: 19462
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/706,691G
; FILING DATE: 19910529
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Drach, John E.
; REGISTRATION NUMBER: 32891
; REFERENCE/DOCKET NUMBER: M4922
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 215-941-6067
; TELEFAX: 215-941-6067
; INFORMATION FOR SEQ ID NO: 9:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 269 amino acids
; TYPE: AMINO ACID
; TOPOLOGY: unknown
; MOLECULE TYPE: protein
; HYPOTHEICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
```

```
; ORGANISM: Serine Protease
; STRAIN: Bacillus lentus DSM 5843
; IMMEDIATE SOURCE:
; CLONE: A188P, V193M
US-07-706-691G-9

Query Match          5.7%; Score 6; DB 1; Length 269;
Best Local Similarity 100.0%; Pred. No. 1.6e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 19 SGVKYA 24
   |||||
Db 24 SGVKYA 29

RESULT 32
US-07-706-691G-10
; Sequence 10, Application US/07706691G
; Patent No. 5340735
; GENERAL INFORMATION:
; APPLICANT: Goddette, Dean W
; TITLE OF INVENTION: Third Generation Protease Mutants
; NUMBER OF SEQUENCES: 104
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Henkel Corporation
; STREET: 140 Germantown Pike, Suite 150
; CITY: Plymouth Meeting
; STATE: Pennsylvania
; COUNTRY: USA
; ZIP: 19462
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/706,691G
; FILING DATE: 19910529
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Drach, John E.
; REGISTRATION NUMBER: 32891
; REFERENCE/DOCKET NUMBER: M4922
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 215-832-2215
; TELEFAX: 215-941-6067
; INFORMATION FOR SEQ ID NO: 10:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 269 amino acids
; TYPE: AMINO ACID
; TOPOLOGY: unknown
; MOLECULE TYPE: protein
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: Serine Protease
; STRAIN: Bacillus lentus DSM 5843
; IMMEDIATE SOURCE:
; CLONE: A188P
US-07-706-691G-10

Query Match          5.7%; Score 6; DB 1; Length 269;
Best Local Similarity 100.0%; Pred. No. 1.6e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 19 SGVKYA 24
   |||||
Db 24 SGVKYA 29

RESULT 33
US-07-706-691G-11
; Sequence 11, Application US/07706691G
```

```
; Patent No. 5340735
; GENERAL INFORMATION:
; APPLICANT: Goddette, Dean W
; TITLE OF INVENTION: Third Generation Protease Mutants
; NUMBER OF SEQUENCES: 104
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Henkel Corporation
; STREET: 140 Germantown Pike, Suite 150
; CITY: Plymouth Meeting
; STATE: Pennsylvania
; COUNTRY: USA
; ZIP: 19462
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/706,691G
; FILING DATE: 19910529
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Drach, John E.
; REGISTRATION NUMBER: 32891
; REFERENCE/DOCKET NUMBER: M4922
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 215-832-2215
; TELEFAX: 215-941-6067
; INFORMATION FOR SEQ ID NO: 11:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 269 amino acids
; TYPE: AMINO ACID
; TOPOLOGY: unknown
; MOLECULE TYPE: protein
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: Serine Protease
; STRAIN: Bacillus lentus DSM 5843
; IMMEDIATE SOURCE:
; CLONE: S3T, V41, A188P, V193M
US-07-706-691G-11

Query Match          5.7%; Score 6; DB 1; Length 269;
Best Local Similarity 100.0%; Pred. No. 1.6e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 19 SGVKYA 24
   |||||
Db 24 SGVKYA 29

RESULT 34
US-07-706-691G-12
; Sequence 12, Application US/07706691G
; Patent No. 5340735
; GENERAL INFORMATION:
; APPLICANT: Goddette, Dean W
; TITLE OF INVENTION: Third Generation Protease Mutants
; NUMBER OF SEQUENCES: 104
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Henkel Corporation
; STREET: 140 Germantown Pike, Suite 150
; CITY: Plymouth Meeting
; STATE: Pennsylvania
; COUNTRY: USA
; ZIP: 19462
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
```

```

; APPLICATION NUMBER: US/07/706,691G
; FILING DATE: 19910529
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
;   NAME: Drach, John E.
;   REGISTRATION NUMBER: 32891
;   REFERENCE/DOCKET NUMBER: M4922
; TELECOMMUNICATION INFORMATION:
;   TELEPHONE: 215-832-2215
;   TELEFAX: 215-941-6067
; INFORMATION FOR SEQ ID NO: 12:
;   SEQUENCE CHARACTERISTICS:
;     LENGTH: 269 amino acids
;     TYPE: AMINO ACID
;     TOPOLOGY: unknown
;     MOLECULE TYPE: protein
;     HYPOTHETICAL: NO
;     ANTI-SENSE: NO
;     ORIGINAL SOURCE:
;       ORGANISM: Serine Protease
;       STRAIN: Bacillus lentus DSM 5843
;     IMMEDIATE SOURCE:
;       CLONE: V193M
; US-07-706-691G-12

```

```

Query Match      5.7%; Score 6; DB 1; Length 269;
Best Local Similarity 100.0%; Pred. No. 1.6e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

Qy 19 SGVKVA 24
    |||||
Db 24 SGVKVA 29

```

```

RESULT 35
US-07-706-691G-13
; Sequence 13, Application US/07706691G
; Patent No. 5340735
; GENERAL INFORMATION:
;   APPLICANT: Goddette, Dean W
;   TITLE OF INVENTION: Third Generation Protease Mutants
;   NUMBER OF SEQUENCES: 104
;   CORRESPONDENCE ADDRESS:
;     ADDRESSEE: Henkel Corporation
;     STREET: 140 Germantown Pike, Suite 150
;     CITY: Plymouth Meeting
;     STATE: Pennsylvania
;     COUNTRY: USA
;     ZIP: 19462
; COMPUTER READABLE FORM:
;   MEDIUM TYPE: Floppy disk
;   COMPUTER: IBM PC compatible
;   OPERATING SYSTEM: PC-DOS/MS-DOS
;   SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
;   APPLICATION NUMBER: US/07/706,691G
;   FILING DATE: 19910529
;   CLASSIFICATION: 435
;   ATTORNEY/AGENT INFORMATION:
;     NAME: Drach, John E.
;     REGISTRATION NUMBER: 32891
;     REFERENCE/DOCKET NUMBER: M4922
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE: 215-832-2215
;       TELEFAX: 215-941-6067
; INFORMATION FOR SEQ ID NO: 13:
;   SEQUENCE CHARACTERISTICS:
;     LENGTH: 269 amino acids
;     TYPE: AMINO ACID
;     TOPOLOGY: unknown
;     MOLECULE TYPE: protein
;     HYPOTHETICAL: NO
;     ANTI-SENSE: NO

```

```

; ORIGINAL SOURCE:
;   ORGANISM: Serine Protease
;   STRAIN: Bacillus lentus DSM 5843
;   IMMEDIATE SOURCE:
;     CLONE: S104T
; US-07-706-691G-13

```

```

Query Match      5.7%; Score 6; DB 1; Length 269;
Best Local Similarity 100.0%; Pred. No. 1.6e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

Qy 19 SGVKVA 24
    |||||
Db 24 SGVKVA 29

```

```

RESULT 36
US-07-706-691G-14
; Sequence 14, Application US/07706691G
; Patent No. 5340735
; GENERAL INFORMATION:
;   APPLICANT: Goddette, Dean W
;   TITLE OF INVENTION: Third Generation Protease Mutants
;   NUMBER OF SEQUENCES: 104
;   CORRESPONDENCE ADDRESS:
;     ADDRESSEE: Henkel Corporation
;     STREET: 140 Germantown Pike, Suite 150
;     CITY: Plymouth Meeting
;     STATE: Pennsylvania
;     COUNTRY: USA
;     ZIP: 19462
; COMPUTER READABLE FORM:
;   MEDIUM TYPE: Floppy disk
;   COMPUTER: IBM PC compatible
;   OPERATING SYSTEM: PC-DOS/MS-DOS
;   SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
;   APPLICATION NUMBER: US/07/706,691G
;   FILING DATE: 19910529
;   CLASSIFICATION: 435
;   ATTORNEY/AGENT INFORMATION:
;     NAME: Drach, John E.
;     REGISTRATION NUMBER: 32891
;     REFERENCE/DOCKET NUMBER: M4922
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE: 215-832-2215
;       TELEFAX: 215-941-6067
; INFORMATION FOR SEQ ID NO: 14:
;   SEQUENCE CHARACTERISTICS:
;     LENGTH: 269 amino acids
;     TYPE: AMINO ACID
;     TOPOLOGY: unknown
;     MOLECULE TYPE: protein
;     HYPOTHETICAL: NO
;     ANTI-SENSE: NO
;     ORIGINAL SOURCE:
;       ORGANISM: Serine Protease
;       STRAIN: Bacillus lentus DSM 5843
;     IMMEDIATE SOURCE:
;       CLONE: T69V
; US-07-706-691G-14

```

```

Query Match      5.7%; Score 6; DB 1; Length 269;
Best Local Similarity 100.0%; Pred. No. 1.6e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

Qy 19 SGVKVA 24
    |||||
Db 24 SGVKVA 29

```

```

RESULT 37
US-07-706-691G-15

```

```
Sequence 15, Application US/07706691G
Patent No. 5340735
GENERAL INFORMATION:
APPLICANT: Goddette, Dean W
TITLE OF INVENTION: Third Generation Protease Mutants
NUMBER OF SEQUENCES: 104
CORRESPONDENCE ADDRESS:
ADDRESSEE: Henkel Corporation
STREET: 140 Germantown Pike, Suite 150
CITY: Plymouth Meeting
STATE: Pennsylvania
COUNTRY: USA
ZIP: 19462
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/706.691G
CLASSIFICATION: 435
FILING DATE: 19910529
ATTORNEY/AGENT INFORMATION:
NAME: Drach, John E.
REGISTRATION NUMBER: 32891
REFERENCE/DOCKET NUMBER: M4922
TELECOMMUNICATION INFORMATION:
TELEPHONE: 215-832-2215
TELEFAX: 215-941-6067
INFORMATION FOR SEQ ID NO: 15:
SEQUENCE CHARACTERISTICS:
LENGTH: 269 amino acids
TYPE: AMINO ACID
TOPOLOGY: unknown
MOLECULE TYPE: protein
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: Serine Protease
STRAIN: Bacillus lentus DSM 5843
IMMEDIATE SOURCE:
CLONE: V41, A188P, V193M
US-07-706-691G-15

Query Match 5.7%; Score 6; DB 1; Length 269;
Best Local Similarity 100.0%; Pred. No. 1.6e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 19 SGVKVA 24
DB 24 SGVKVA 29

RESULT 38
US-07-706-691G-16
Sequence 16, Application US/07706691G
Patent No. 5340735
GENERAL INFORMATION:
APPLICANT: Goddette, Dean W
TITLE OF INVENTION: Third Generation Protease Mutants
NUMBER OF SEQUENCES: 104
CORRESPONDENCE ADDRESS:
ADDRESSEE: Henkel Corporation
STREET: 140 Germantown Pike, Suite 150
CITY: Plymouth Meeting
STATE: Pennsylvania
COUNTRY: USA
ZIP: 19462
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
```

```
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/706.691G
FILING DATE: 19910529
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Drach, John E.
REGISTRATION NUMBER: 32891
REFERENCE/DOCKET NUMBER: M4922
TELECOMMUNICATION INFORMATION:
TELEPHONE: 215-832-2215
TELEFAX: 215-941-6067
INFORMATION FOR SEQ ID NO: 16:
SEQUENCE CHARACTERISTICS:
LENGTH: 269 amino acids
TYPE: AMINO ACID
TOPOLOGY: unknown
MOLECULE TYPE: protein
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: Serine Protease
STRAIN: Bacillus lentus DSM 5843
IMMEDIATE SOURCE:
CLONE: A224V
US-07-706-691G-16

Query Match 5.7%; Score 6; DB 1; Length 269;
Best Local Similarity 100.0%; Pred. No. 1.6e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 19 SGVKVA 24
DB 24 SGVKVA 29

RESULT 39
US-07-706-691G-17
Sequence 17, Application US/07706691G
Patent No. 5340735
GENERAL INFORMATION:
APPLICANT: Goddette, Dean W
TITLE OF INVENTION: Third Generation Protease Mutants
NUMBER OF SEQUENCES: 104
CORRESPONDENCE ADDRESS:
ADDRESSEE: Henkel Corporation
STREET: 140 Germantown Pike, Suite 150
CITY: Plymouth Meeting
STATE: Pennsylvania
COUNTRY: USA
ZIP: 19462
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/706.691G
FILING DATE: 19910529
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Drach, John E.
REGISTRATION NUMBER: 32891
REFERENCE/DOCKET NUMBER: M4922
TELECOMMUNICATION INFORMATION:
TELEPHONE: 215-832-2215
TELEFAX: 215-941-6067
INFORMATION FOR SEQ ID NO: 17:
SEQUENCE CHARACTERISTICS:
LENGTH: 269 amino acids
TYPE: AMINO ACID
TOPOLOGY: unknown
MOLECULE TYPE: protein
HYPOTHETICAL: NO
```

ANTI-SENSE: NO  
ORIGINAL SOURCE:  
ORGANISM: Serine Protease  
STRAIN: Bacillus lentus DSM 5843  
IMMEDIATE SOURCE:  
CLONE: V1991  
US-07-706-691G-17

Job time : 45.0164 secs

Query Match 5.7%; Score 6; DB 1; Length 269;  
Best Local Similarity 100.0%; Pred. No. 1.6e+02;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 19 SGVKA 24  
|||  
Db 24 SGVKA 29

RESULT 40  
US-07-706-691G-18  
Sequence 18, Application US/07706691G  
Patent No. 5340735  
GENERAL INFORMATION:  
APPLICANT: Goddette, Dean W  
TITLE OF INVENTION: Third Generation Protease Mutants  
NUMBER OF SEQUENCES: 104  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Henkel Corporation  
STREET: 140 Germantown Pike, Suite 150  
CITY: Plymouth Meeting  
STATE: Pennsylvania  
COUNTRY: USA  
ZIP: 19462  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/07/706,691G  
FILING DATE: 19910529  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Drach, John E.  
REGISTRATION NUMBER: 32891  
REFERENCE/DOCKET NUMBER: M4922  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 215-832-2215  
TELEFAX: 215-941-6067  
INFORMATION FOR SEQ ID NO: 18:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 269 amino acids  
TYPE: AMINO ACID  
TOPOLOGY: unknown  
MOLECULE TYPE: Protein  
HYPOTHETICAL: NO  
ANTI-SENSE: NO  
ORIGINAL SOURCE:  
ORGANISM: Serine Protease  
STRAIN: Bacillus lentus DSM 5843  
IMMEDIATE SOURCE:  
CLONE: V41  
US-07-706-691G-18

Query Match 5.7%; Score 6; DB 1; Length 269;  
Best Local Similarity 100.0%; Pred. No. 1.6e+02;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 19 SGVKA 24  
|||  
Db 24 SGVKA 29

Search completed: April 9, 2003, 13:16:26

GenCore version 5.1.4 p5.4578  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 9, 2003, 13:15:43 ; Search time 39.0984 Seconds  
(without alignments)  
165.747 Million cell updates/sec

Title: US-10-020-441-2

Perfect score: 106  
1 VSFVLRNGDQGIHQPSG.....YGRDDNDPRSPKNGILKK 106

Scoring table: OLIGO  
Gapop 60.0 , Gapext 60.0

Searched: 248812 seqs, 61136040 residues

Word size: 0

Total number of hits satisfying chosen parameters: 248812

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Listing first 1000 summaries

Database:

Published Applications\_AA:\*

- 1: /cgn2\_6/ptodata/2/pubpaa/US08\_NEW\_PUB.pep:\*
- 2: /cgn2\_6/ptodata/2/pubpaa/PCR\_NEW\_PUB.pep:\*
- 3: /cgn2\_6/ptodata/2/pubpaa/US06\_NEW\_PUB.pep:\*
- 4: /cgn2\_6/ptodata/2/pubpaa/US06\_PUBCOMB.pep:\*
- 5: /cgn2\_6/ptodata/2/pubpaa/US07\_NEW\_PUB.pep:\*
- 6: /cgn2\_6/ptodata/2/pubpaa/US07\_PUBCOMB.pep:\*
- 7: /cgn2\_6/ptodata/2/pubpaa/PCUTS\_PUBCOMB.pep:\*
- 8: /cgn2\_6/ptodata/2/pubpaa/US08\_PUBCOMB.pep:\*
- 9: /cgn2\_6/ptodata/2/pubpaa/US08\_NEW\_PUB.pep:\*
- 10: /cgn2\_6/ptodata/2/pubpaa/US05\_PUBCOMB.pep:\*
- 11: /cgn2\_6/ptodata/2/pubpaa/US10\_NEW\_PUB.pep:\*
- 12: /cgn2\_6/ptodata/2/pubpaa/US10\_PUBCOMB.pep:\*
- 13: /cgn2\_6/ptodata/2/pubpaa/US60\_NEW\_PUB.pep:\*
- 14: /cgn2\_6/ptodata/2/pubpaa/US60\_PUBCOMB.pep:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	106	100.0	106	US-10-020-441-2	Sequence 2, Appli
2	106	15.1	16	US-10-020-441-3	Sequence 3, Appli
3	7	6.6	359	US-10-016-668-4	Sequence 4, Appli
4	7	6.6	1737	US-09-808-602-83	Sequence 83, Appli
5	7	6.6	2724	US-09-808-602-13	Sequence 13, Appli
6	7	6.6	2733	US-09-808-602-8	Sequence 8, Appli
7	7	6.6	2764	US-09-808-602-80	Sequence 80, Appli
8	7	6.6	2765	US-09-808-602-84	Sequence 84, Appli
9	7	6.6	2802	US-09-808-602-81	Sequence 81, Appli
10	6	5.7	21	US-09-027-556-6	Sequence 6, Appli
11	6	5.7	30	US-09-864-761-47769	Sequence 47769, A
12	6	5.7	34	US-09-864-761-38283	Sequence 1791, Ap
13	6	5.7	75	US-09-796-692-1791	Sequence 2055, Ap
14	6	5.7	75	US-09-796-692-2055	Sequence 521, App
15	6	5.7	119	US-09-925-302-521	Sequence 36308, A
16	6	5.7	128	US-09-864-761-36308	Sequence 5580, Ap
17	6	5.7	144	US-09-738-626-5580	Sequence 34045, A
18	6	5.7	146	US-09-864-761-34045	Sequence 158, App
19	6	5.7	170	US-09-975-719-158	

20	6	5.7	177	US-09-815-242-5164	Sequence 5164, Ap
21	6	5.7	196	US-09-811-284-207	Sequence 207, App
22	6	5.7	218	US-09-738-626-3636	Sequence 3636, Ap
23	6	5.7	220	US-09-938-803-4	Sequence 4, Appli
24	6	5.7	221	US-09-881-752A-60	Sequence 60, Appli
25	6	5.7	255	US-09-864-761-35704	Sequence 35704, A
26	6	5.7	263	US-09-815-242-14092	Sequence 14092, A
27	6	5.7	266	US-09-808-665A-6	Sequence 6, Appli
28	6	5.7	269	US-09-975-139-1	Sequence 1, Appli
29	6	5.7	269	US-09-976-414-8	Sequence 8, Appli
30	6	5.7	269	US-09-813-408-3	Sequence 3, Appli
31	6	5.7	269	US-09-813-408-4	Sequence 4, Appli
32	6	5.7	269	US-09-813-408-5	Sequence 5, Appli
33	6	5.7	269	US-09-837-235-16	Sequence 16, Appli
34	6	5.7	269	US-09-060-854B-16	Sequence 6, Appli
35	6	5.7	269	US-10-075-907-1	Sequence 1, Appli
36	6	5.7	269	US-10-075-895-1	Sequence 1, Appli
37	6	5.7	271	US-09-813-408-2	Sequence 2, Appli
38	6	5.7	285	US-09-815-242-10241	Sequence 10241, A
39	6	5.7	325	US-09-738-626-6072	Sequence 6072, Ap
40	6	5.7	326	US-09-252-088-35	Sequence 35, Appli
41	6	5.7	334	US-09-454-279-21	Sequence 21, Appli
42	6	5.7	371	US-09-738-626-5312	Sequence 5312, Ap
43	6	5.7	377	US-09-813-408-1	Sequence 1, Appli
44	6	5.7	380	US-09-824-893A-261	Sequence 261, App
45	6	5.7	394	US-10-029-654-14	Sequence 14, Appli
46	6	5.7	417	US-09-972-268-25	Sequence 25, Appli
47	6	5.7	453	US-09-925-301-1286	Sequence 1286, Ap
48	6	5.7	471	US-09-120-044-3	Sequence 3, Appli
49	6	5.7	471	US-09-120-044-4	Sequence 4, Appli
50	6	5.7	537	US-09-815-242-13463	Sequence 13463, A
51	6	5.7	543	US-10-033-297-143	Sequence 143, App
52	6	5.7	543	US-09-940-244-143	Sequence 143, App
53	6	5.7	655	US-10-054-683-29	Sequence 29, Appli
54	6	5.7	655	US-10-001-632A-2	Sequence 2, Appli
55	6	5.7	737	US-09-808-665A-2	Sequence 2, Appli
56	6	5.7	737	US-09-808-665A-4	Sequence 4, Appli
57	6	5.7	934	US-08-837-459-19	Sequence 19, Appli
58	6	5.7	2165	US-09-923-070A-29	Sequence 29, Appli
59	6	5.7	2165	US-09-923-070A-30	Sequence 30, Appli
60	6	5.7	2165	US-09-161-122-29	Sequence 29, Appli
61	6	5.7	2165	US-09-161-122-30	Sequence 30, Appli
62	6	5.7	5795	US-09-815-242-12610	Sequence 12610, A
63	6	4.7	9	US-10-106-487-16	Sequence 16, Appli
64	5	4.7	9	US-09-909-460-66	Sequence 66, Appli
65	5	4.7	9	US-09-277-074-3	Sequence 3, Appli
66	5	4.7	9	US-09-894-018-326	Sequence 326, App
67	5	4.7	9	US-09-792-480-16	Sequence 16, Appli
68	5	4.7	10	US-09-277-074-27	Sequence 27, Appli
69	5	4.7	11	US-09-017-743C-122	Sequence 122, App
70	5	4.7	15	US-10-126-474A-4	Sequence 4, Appli
71	5	4.7	17	US-09-864-761-48339	Sequence 48339, A
72	5	4.7	18	US-09-852-797-113	Sequence 113, App
73	5	4.7	18	US-09-962-445-5	Sequence 5, Appli
74	5	4.7	18	US-09-853-161-113	Sequence 113, App
75	5	4.7	18	US-09-852-659A-113	Sequence 113, App
76	5	4.7	19	US-10-071-512A-6	Sequence 6, Appli
77	5	4.7	20	US-09-902-941-1969	Sequence 1969, Ap
78	5	4.7	20	US-09-902-941-1970	Sequence 1970, Ap
79	5	4.7	20	US-09-902-941-1971	Sequence 1971, Ap
80	5	4.7	20	US-10-017-754-1969	Sequence 1969, Ap
81	5	4.7	20	US-10-017-754-1970	Sequence 1970, Ap
82	5	4.7	20	US-10-017-754-1971	Sequence 1971, Ap
83	5	4.7	24	US-09-963-602-493	Sequence 493, App
84	5	4.7	26	US-09-864-761-48446	Sequence 48446, A
85	5	4.7	26	US-10-124-557-18	Sequence 18, Appli
86	5	4.7	27	US-09-939-528-335	Sequence 335, App
87	5	4.7	31	US-09-812-528-9	Sequence 9, Appli
88	5	4.7	33	US-10-081-816-112	Sequence 112, App
89	5	4.7	33	US-10-007-280A-1237	Sequence 237, App
90	5	4.7	33	US-09-864-761-38816	Sequence 38816, A
91	5	4.7	35	US-09-864-761-41273	Sequence 41273, A
92	5	4.7	35	US-09-949-780-8	Sequence 8, Appli





239	5	4.7	139	12	US-10-096-241-33	Sequence 33, App1	312	5	4.7	159	9	US-10-176-485-540	Sequence 540, App
240	5	4.7	140	9	US-09-974-879-331	Sequence 331, App	313	5	4.7	159	9	US-10-176-487-540	Sequence 540, App
241	5	4.7	142	9	US-10-079-623-367	Sequence 367, App	314	5	4.7	159	9	US-10-176-493-540	Sequence 540, App
242	5	4.7	142	10	US-09-380-882-8	Sequence 8, App1	315	5	4.7	159	9	US-10-176-756-540	Sequence 540, App
243	5	4.7	142	10	US-09-888-035-8	Sequence 8, App1	316	5	4.7	159	9	US-10-176-911-540	Sequence 540, App
244	5	4.7	142	10	US-09-862-027-31	Sequence 31, App1	317	5	4.7	159	9	US-10-176-919-540	Sequence 540, App
245	5	4.7	143	9	US-10-101-464A-671	Sequence 671, App	318	5	4.7	159	9	US-10-176-925-540	Sequence 540, App
246	5	4.7	144	9	US-10-076-785-50	Sequence 50, App1	319	5	4.7	159	9	US-10-176-978-540	Sequence 540, App
247	5	4.7	144	10	US-09-815-242-5160	Sequence 5160, App	320	5	4.7	159	9	US-10-179-510-540	Sequence 540, App
248	5	4.7	144	10	US-09-894-018-129	Sequence 129, App	321	5	4.7	159	9	US-10-180-543-540	Sequence 540, App
249	5	4.7	145	9	US-09-764-904-48	Sequence 48, App1	322	5	4.7	159	9	US-10-180-544-540	Sequence 540, App
250	5	4.7	145	9	US-10-091-548-48	Sequence 48, App1	323	5	4.7	159	9	US-10-180-546-540	Sequence 540, App
251	5	4.7	145	10	US-09-764-860-562	Sequence 562, App	324	5	4.7	159	9	US-10-180-547-540	Sequence 540, App
252	5	4.7	146	10	US-09-864-761-41738	Sequence 41738, A	325	5	4.7	159	9	US-10-180-549-540	Sequence 540, App
253	5	4.7	146	10	US-09-893-737-42	Sequence 42, App1	326	5	4.7	159	9	US-10-180-555-540	Sequence 540, App
254	5	4.7	147	10	US-09-894-018-131	Sequence 131, App	327	5	4.7	159	9	US-10-180-559-540	Sequence 540, App
255	5	4.7	148	9	US-10-090-569-14	Sequence 14, App1	328	5	4.7	159	9	US-10-181-000-540	Sequence 540, App
256	5	4.7	148	10	US-09-016-869B-35	Sequence 35, App1	329	5	4.7	159	9	US-10-183-010-540	Sequence 540, App
257	5	4.7	148	10	US-09-894-018-127	Sequence 127, App	330	5	4.7	159	9	US-10-183-012-540	Sequence 540, App
258	5	4.7	150	10	US-09-529-063-22	Sequence 22, App1	331	5	4.7	159	9	US-10-184-514-540	Sequence 540, App
259	5	4.7	150	12	US-10-021-509-2	Sequence 2, App1	332	5	4.7	159	9	US-10-184-623-540	Sequence 540, App
260	5	4.7	151	10	US-09-883-985-10	Sequence 10, App1	333	5	4.7	159	9	US-10-184-635-540	Sequence 540, App
261	5	4.7	152	10	US-09-864-761-47290	Sequence 47290, A	334	5	4.7	159	9	US-10-184-637-540	Sequence 540, App
262	5	4.7	154	9	US-09-738-626-6242	Sequence 6242, Ap	335	5	4.7	159	9	US-10-184-646-540	Sequence 540, App
263	5	4.7	154	9	US-10-101-464A-99	Sequence 99, App1	336	5	4.7	159	9	US-10-184-647-540	Sequence 540, App
264	5	4.7	154	9	US-10-101-464A-768	Sequence 768, App	337	5	4.7	159	9	US-10-184-652-540	Sequence 540, App
265	5	4.7	155	9	US-09-712-363-155	Sequence 155, App	338	5	4.7	159	9	US-10-187-594-540	Sequence 540, App
266	5	4.7	156	9	US-09-802-941-1908	Sequence 1908, Ap	339	5	4.7	159	9	US-10-187-596-540	Sequence 540, App
267	5	4.7	156	9	US-09-849-626-1908	Sequence 1908, Ap	340	5	4.7	159	9	US-10-187-745-540	Sequence 540, App
268	5	4.7	156	9	US-10-017-754-1908	Sequence 1908, Ap	341	5	4.7	159	9	US-10-187-885-540	Sequence 540, App
269	5	4.7	156	10	US-09-016-750C-2	Sequence 2, App1	342	5	4.7	159	9	US-10-187-886-540	Sequence 540, App
270	5	4.7	156	10	US-09-864-761-43022	Sequence 43022, A	343	5	4.7	159	9	US-10-187-886-540	Sequence 540, App
271	5	4.7	156	10	US-09-016-869B-2	Sequence 2, App1	344	5	4.7	159	9	US-10-176-751-540	Sequence 540, App
272	5	4.7	157	10	US-09-016-750C-12	Sequence 12, App1	345	5	4.7	159	9	US-10-176-760-540	Sequence 540, App
273	5	4.7	157	10	US-09-016-869B-12	Sequence 12, App1	346	5	4.7	159	9	US-10-176-990-540	Sequence 540, App
274	5	4.7	159	9	US-09-902-941-1937	Sequence 1937, Ap	347	5	4.7	159	9	US-10-180-541-540	Sequence 540, App
275	5	4.7	159	9	US-10-176-559-540	Sequence 540, App	348	5	4.7	159	9	US-10-180-542-540	Sequence 540, App
276	5	4.7	159	9	US-10-176-758-540	Sequence 540, App	349	5	4.7	159	9	US-10-180-548-540	Sequence 540, App
277	5	4.7	159	9	US-10-175-737-540	Sequence 540, App	350	5	4.7	159	9	US-10-180-551-540	Sequence 540, App
278	5	4.7	159	9	US-10-173-706-540	Sequence 540, App	351	5	4.7	159	9	US-10-180-998-540	Sequence 540, App
279	5	4.7	159	9	US-10-175-738-540	Sequence 540, App	352	5	4.7	159	9	US-10-180-999-540	Sequence 540, App
280	5	4.7	159	9	US-10-175-752-540	Sequence 540, App	353	5	4.7	159	9	US-10-183-013-540	Sequence 540, App
281	5	4.7	159	9	US-10-176-482-540	Sequence 540, App	354	5	4.7	159	9	US-10-184-612-540	Sequence 540, App
282	5	4.7	159	9	US-10-176-757-540	Sequence 540, App	355	5	4.7	159	9	US-10-184-616-540	Sequence 540, App
283	5	4.7	159	9	US-10-176-913-540	Sequence 540, App	356	5	4.7	159	9	US-10-184-617-540	Sequence 540, App
284	5	4.7	159	9	US-10-180-552-540	Sequence 540, App	357	5	4.7	159	9	US-10-184-622-540	Sequence 540, App
285	5	4.7	159	9	US-10-180-557-540	Sequence 540, App	358	5	4.7	159	9	US-10-184-628-540	Sequence 540, App
286	5	4.7	159	9	US-10-173-700-540	Sequence 540, App	359	5	4.7	159	9	US-10-184-629-540	Sequence 540, App
287	5	4.7	159	9	US-10-174-572-540	Sequence 540, App	360	5	4.7	159	9	US-10-184-630-540	Sequence 540, App
288	5	4.7	159	9	US-10-174-579-540	Sequence 540, App	361	5	4.7	159	9	US-10-184-631-540	Sequence 540, App
289	5	4.7	159	9	US-10-174-582-540	Sequence 540, App	362	5	4.7	159	9	US-10-184-632-540	Sequence 540, App
290	5	4.7	159	9	US-10-174-588-540	Sequence 540, App	363	5	4.7	159	9	US-10-184-636-540	Sequence 540, App
291	5	4.7	159	9	US-10-175-739-540	Sequence 540, App	364	5	4.7	159	9	US-10-184-640-540	Sequence 540, App
292	5	4.7	159	9	US-10-175-740-540	Sequence 540, App	365	5	4.7	159	9	US-10-184-650-540	Sequence 540, App
293	5	4.7	159	9	US-10-175-743-540	Sequence 540, App	366	5	4.7	159	9	US-10-184-651-540	Sequence 540, App
294	5	4.7	159	9	US-10-176-488-540	Sequence 540, App	367	5	4.7	159	9	US-10-184-658-540	Sequence 540, App
295	5	4.7	159	9	US-10-176-492-540	Sequence 540, App	368	5	4.7	159	9	US-10-187-597-540	Sequence 540, App
296	5	4.7	159	9	US-10-176-717-540	Sequence 540, App	369	5	4.7	159	9	US-10-187-598-540	Sequence 540, App
297	5	4.7	159	9	US-10-176-750-540	Sequence 540, App	370	5	4.7	159	9	US-10-187-600-540	Sequence 540, App
298	5	4.7	159	9	US-10-176-985-540	Sequence 540, App	371	5	4.7	159	9	US-10-187-601-540	Sequence 540, App
299	5	4.7	159	9	US-10-176-987-540	Sequence 540, App	372	5	4.7	159	9	US-10-187-602-540	Sequence 540, App
300	5	4.7	159	9	US-10-176-991-540	Sequence 540, App	373	5	4.7	159	9	US-10-187-603-540	Sequence 540, App
301	5	4.7	159	9	US-10-176-992-540	Sequence 540, App	374	5	4.7	159	9	US-10-187-741-540	Sequence 540, App
302	5	4.7	159	9	US-10-176-993-540	Sequence 540, App	375	5	4.7	159	9	US-10-187-743-540	Sequence 540, App
303	5	4.7	159	9	US-10-184-658-540	Sequence 540, App	376	5	4.7	159	9	US-10-187-746-540	Sequence 540, App
304	5	4.7	159	9	US-10-173-695-540	Sequence 540, App	377	5	4.7	159	9	US-10-187-747-540	Sequence 540, App
305	5	4.7	159	9	US-10-173-697-540	Sequence 540, App	378	5	4.7	159	9	US-10-187-751-540	Sequence 540, App
306	5	4.7	159	9	US-10-173-705-540	Sequence 540, App	379	5	4.7	159	9	US-10-187-753-540	Sequence 540, App
307	5	4.7	159	9	US-10-174-576-540	Sequence 540, App	380	5	4.7	159	9	US-10-187-754-540	Sequence 540, App
308	5	4.7	159	9	US-10-174-585-540	Sequence 540, App	381	5	4.7	159	9	US-10-187-757-540	Sequence 540, App
309	5	4.7	159	9	US-10-174-586-540	Sequence 540, App	382	5	4.7	159	9	US-10-187-884-540	Sequence 540, App
310	5	4.7	159	9	US-10-175-747-540	Sequence 540, App	383	5	4.7	159	9	US-10-188-767-540	Sequence 540, App
311	5	4.7	159	9	US-10-176-481-540	Sequence 540, App	384	5	4.7	159	9	US-10-188-769-540	Sequence 540, App

385	5	4.7	159	9	US-10-188-770-540	Sequence 540, App	458	5	4.7	159	9	US-10-197-708-540	Sequence 540, App
386	5	4.7	159	9	US-10-188-773-540	Sequence 540, App	459	5	4.7	159	9	US-10-198-764-540	Sequence 540, App
387	5	4.7	159	9	US-10-188-781-540	Sequence 540, App	460	5	4.7	159	9	US-10-198-765-540	Sequence 540, App
388	5	4.7	159	9	US-10-194-361-540	Sequence 540, App	461	5	4.7	159	9	US-10-198-768-540	Sequence 540, App
389	5	4.7	159	9	US-10-194-423-540	Sequence 540, App	462	5	4.7	159	9	US-10-198-769-540	Sequence 540, App
390	5	4.7	159	9	US-10-195-897-540	Sequence 540, App	463	5	4.7	159	9	US-10-199-305-540	Sequence 540, App
391	5	4.7	159	9	US-10-195-901-540	Sequence 540, App	464	5	4.7	159	9	US-10-199-306-540	Sequence 540, App
392	5	4.7	159	9	US-10-196-756-540	Sequence 540, App	465	5	4.7	159	9	US-10-199-310-540	Sequence 540, App
393	5	4.7	159	9	US-10-173-708-540	Sequence 540, App	466	5	4.7	159	9	US-10-199-311-540	Sequence 540, App
394	5	4.7	159	9	US-10-176-479-540	Sequence 540, App	467	5	4.7	159	9	US-10-199-314-540	Sequence 540, App
395	5	4.7	159	9	US-10-176-748-540	Sequence 540, App	468	5	4.7	159	9	US-10-199-317-540	Sequence 540, App
396	5	4.7	159	9	US-10-176-916-540	Sequence 540, App	469	5	4.7	159	9	US-10-199-665-540	Sequence 540, App
397	5	4.7	159	9	US-10-179-507-540	Sequence 540, App	470	5	4.7	159	9	US-10-199-666-540	Sequence 540, App
398	5	4.7	159	9	US-10-179-516-540	Sequence 540, App	471	5	4.7	159	9	US-10-199-669-540	Sequence 540, App
399	5	4.7	159	9	US-10-179-519-540	Sequence 540, App	472	5	4.7	159	9	US-10-201-534-540	Sequence 540, App
400	5	4.7	159	9	US-10-179-525-540	Sequence 540, App	473	5	4.7	159	9	US-10-201-770-540	Sequence 540, App
401	5	4.7	159	9	US-10-180-545-540	Sequence 540, App	474	5	4.7	159	9	US-10-201-855-540	Sequence 540, App
402	5	4.7	159	9	US-10-180-548-540	Sequence 540, App	475	5	4.7	159	9	US-10-201-856-540	Sequence 540, App
403	5	4.7	159	9	US-10-183-006-540	Sequence 540, App	476	5	4.7	159	9	US-10-202-469-540	Sequence 540, App
404	5	4.7	159	9	US-10-183-008-540	Sequence 540, App	477	5	4.7	159	9	US-10-202-470-540	Sequence 540, App
405	5	4.7	159	9	US-10-183-017-540	Sequence 540, App	478	5	4.7	159	9	US-10-202-476-540	Sequence 540, App
406	5	4.7	159	9	US-10-183-019-540	Sequence 540, App	479	5	4.7	159	9	US-10-202-934-540	Sequence 540, App
407	5	4.7	159	9	US-10-184-618-540	Sequence 540, App	480	5	4.7	159	9	US-10-202-935-540	Sequence 540, App
408	5	4.7	159	9	US-10-184-625-540	Sequence 540, App	481	5	4.7	159	9	US-10-202-936-540	Sequence 540, App
409	5	4.7	159	9	US-10-184-626-540	Sequence 540, App	482	5	4.7	159	9	US-10-202-939-540	Sequence 540, App
410	5	4.7	159	9	US-10-184-627-540	Sequence 540, App	483	5	4.7	159	9	US-10-205-504-540	Sequence 540, App
411	5	4.7	159	9	US-10-184-645-540	Sequence 540, App	484	5	4.7	159	9	US-10-205-509-540	Sequence 540, App
412	5	4.7	159	9	US-10-184-654-540	Sequence 540, App	485	5	4.7	159	9	US-10-205-895-540	Sequence 540, App
413	5	4.7	159	9	US-10-184-655-540	Sequence 540, App	486	5	4.7	159	9	US-10-205-	

531	5	4.7	160	9	US-09-849-626-809	Sequence 809, App	604	5	4.7	173	9	US-10-176-978-506	Sequence 506, App
532	5	4.7	160	9	US-10-017-754-809	Sequence 809, App	605	5	4.7	173	9	US-10-179-510-506	Sequence 506, App
533	5	4.7	160	9	US-10-017-754-1940	Sequence 1940, App	606	5	4.7	173	9	US-10-180-543-506	Sequence 506, App
534	5	4.7	161	9	US-10-165-603-30	Sequence 30, App1	607	5	4.7	173	9	US-10-180-544-506	Sequence 506, App
535	5	4.7	162	9	US-10-013-379-28	Sequence 28, App1	608	5	4.7	173	9	US-10-180-546-506	Sequence 506, App
536	5	4.7	163	10	US-09-761-782-2	Sequence 2, App1	609	5	4.7	173	9	US-10-180-547-506	Sequence 506, App
537	5	4.7	163	10	US-09-815-242-10023	Sequence 12, App1	610	5	4.7	173	9	US-10-180-549-506	Sequence 506, App
538	5	4.7	164	9	US-09-920-395A-12	Sequence 5538, App	611	5	4.7	173	9	US-10-180-555-506	Sequence 506, App
539	5	4.7	164	10	US-09-815-242-5538	Sequence 74, App1	612	5	4.7	173	9	US-10-180-559-506	Sequence 506, App
540	5	4.7	164	12	US-10-078-929-74	Sequence 82, App1	613	5	4.7	173	9	US-10-181-000-506	Sequence 506, App
541	5	4.7	164	12	US-10-078-929-82	Sequence 34905, A	614	5	4.7	173	9	US-10-183-010-506	Sequence 506, App
542	5	4.7	165	10	US-09-864-761-34905	Sequence 28, App1	615	5	4.7	173	9	US-10-183-012-506	Sequence 506, App
543	5	4.7	166	9	US-09-908-931B-28	Sequence 841, App	616	5	4.7	173	9	US-10-184-614-506	Sequence 506, App
544	5	4.7	166	9	US-09-764-868-841	Sequence 34, App1	617	5	4.7	173	9	US-10-184-623-506	Sequence 506, App
545	5	4.7	166	9	US-10-084-205-34	Sequence 12222, A	618	5	4.7	173	9	US-10-184-635-506	Sequence 506, App
546	5	4.7	166	10	US-09-815-242-12222	Sequence 12826, A	619	5	4.7	173	9	US-10-184-637-506	Sequence 506, App
547	5	4.7	166	10	US-09-815-242-12828	Sequence 34, App1	620	5	4.7	173	9	US-10-184-646-506	Sequence 506, App
548	5	4.7	166	10	US-09-925-637-34	Sequence 1921, App	621	5	4.7	173	9	US-10-184-647-506	Sequence 506, App
549	5	4.7	167	9	US-09-902-941-1921	Sequence 1921, App	622	5	4.7	173	9	US-10-184-652-506	Sequence 506, App
550	5	4.7	167	9	US-09-849-626-1921	Sequence 1921, App	623	5	4.7	173	9	US-10-187-594-506	Sequence 506, App
551	5	4.7	167	9	US-10-017-754-1921	Sequence 1921, App	624	5	4.7	173	9	US-10-187-596-506	Sequence 506, App
552	5	4.7	167	10	US-09-765-272-138	Sequence 138, App	625	5	4.7	173	9	US-10-187-745-506	Sequence 506, App
553	5	4.7	170	9	US-09-986-480-282	Sequence 282, App	626	5	4.7	173	9	US-10-187-885-506	Sequence 506, App
554	5	4.7	170	9	US-10-101-464A-116	Sequence 116, App1	627	5	4.7	173	9	US-10-187-886-506	Sequence 506, App
555	5	4.7	171	10	US-09-354-453-9	Sequence 9, App1	628	5	4.7	173	9	US-10-187-886-506	Sequence 506, App
556	5	4.7	172	9	US-10-076-785-56	Sequence 56, App1	629	5	4.7	173	9	US-10-176-751-506	Sequence 506, App
557	5	4.7	172	9	US-09-809-391-760	Sequence 760, App	630	5	4.7	173	9	US-10-176-760-506	Sequence 506, App
558	5	4.7	172	10	US-09-354-453-23	Sequence 23, App1	631	5	4.7	173	9	US-10-176-990-506	Sequence 506, App
559	5	4.7	173	9	US-09-764-868-907	Sequence 907, App	632	5	4.7	173	9	US-10-180-541-506	Sequence 506, App
560	5	4.7	173	9	US-10-174-590-506	Sequence 506, App	633	5	4.7	173	9	US-10-180-542-506	Sequence 506, App
561	5	4.7	173	9	US-10-176-758-506	Sequence 506, App	634	5	4.7	173	9	US-10-180-548-506	Sequence 506, App
562	5	4.7	173	9	US-10-175-737-506	Sequence 506, App	635	5	4.7	173	9	US-10-180-551-506	Sequence 506, App
563	5	4.7	173	9	US-10-173-706-506	Sequence 506, App	636	5	4.7	173	9	US-10-180-998-506	Sequence 506, App
564	5	4.7	173	9	US-10-175-738-506	Sequence 506, App	637	5	4.7	173	9	US-10-180-999-506	Sequence 506, App
565	5	4.7	173	9	US-10-175-732-506	Sequence 506, App	638	5	4.7	173	9	US-10-183-013-506	Sequence 506, App
566	5	4.7	173	9	US-10-176-482-506	Sequence 506, App	639	5	4.7	173	9	US-10-184-612-506	Sequence 506, App
567	5	4.7	173	9	US-10-176-757-506	Sequence 506, App	640	5	4.7	173	9	US-10-184-616-506	Sequence 506, App
568	5	4.7	173	9	US-10-176-913-506	Sequence 506, App	641	5	4.7	173	9	US-10-184-617-506	Sequence 506, App
569	5	4.7	173	9	US-10-180-552-506	Sequence 506, App	642	5	4.7	173	9	US-10-184-622-506	Sequence 506, App
570	5	4.7	173	9	US-10-180-557-506	Sequence 506, App	643	5	4.7	173	9	US-10-184-628-506	Sequence 506, App
571	5	4.7	173	9	US-10-173-700-506	Sequence 506, App	644	5	4.7	173	9	US-10-184-629-506	Sequence 506, App
572	5	4.7	173	9	US-10-174-572-506	Sequence 506, App	645	5	4.7	173	9	US-10-184-630-506	Sequence 506, App
573	5	4.7	173	9	US-10-174-579-506	Sequence 506, App	646	5	4.7	173	9	US-10-184-631-506	Sequence 506, App
574	5	4.7	173	9	US-10-174-582-506	Sequence 506, App	647	5	4.7	173	9	US-10-184-632-506	Sequence 506, App
575	5	4.7	173	9	US-10-174-588-506	Sequence 506, App	648	5	4.7	173	9	US-10-184-636-506	Sequence 506, App
576	5	4.7	173	9	US-10-176-739-506	Sequence 506, App	649	5	4.7	173	9	US-10-184-640-506	Sequence 506, App
577	5	4.7	173	9	US-10-175-740-506	Sequence 506, App	650	5	4.7	173	9	US-10-184-650-506	Sequence 506, App
578	5	4.7	173	9	US-10-175-743-506	Sequence 506, App	651	5	4.7	173	9	US-10-184-651-506	Sequence 506, App
579	5	4.7	173	9	US-10-176-488-506	Sequence 506, App	652	5	4.7	173	9	US-10-187-888-506	Sequence 506, App
580	5	4.7	173	9	US-10-176-492-506	Sequence 506, App	653	5	4.7	173	9	US-10-187-597-506	Sequence 506, App
581	5	4.7	173	9	US-10-176-747-506	Sequence 506, App	654	5	4.7	173	9	US-10-187-598-506	Sequence 506, App
582	5	4.7	173	9	US-10-176-750-506	Sequence 506, App	655	5	4.7	173	9	US-10-187-599-506	Sequence 506, App
583	5	4.7	173	9	US-10-176-985-506	Sequence 506, App	656	5	4.7	173	9	US-10-187-601-506	Sequence 506, App
584	5	4.7	173	9	US-10-176-987-506	Sequence 506, App	657	5	4.7	173	9	US-10-187-602-506	Sequence 506, App
585	5	4.7	173	9	US-10-176-991-506	Sequence 506, App	658	5	4.7	173	9	US-10-187-751-506	Sequence 506, App
586	5	4.7	173	9	US-10-176-992-506	Sequence 506, App	659	5	4.7	173	9	US-10-187-741-506	Sequence 506, App
587	5	4.7	173	9	US-10-176-993-506	Sequence 506, App	660	5	4.7	173	9	US-10-187-743-506	Sequence 506, App
588	5	4.7	173	9	US-10-184-658-506	Sequence 506, App	661	5	4.7	173	9	US-10-187-746-506	Sequence 506, App
589	5	4.7	173	9	US-10-173-695-506	Sequence 506, App	662	5	4.7	173	9	US-10-187-747-506	Sequence 506, App
590	5	4.7	173	9	US-10-173-697-506	Sequence 506, App	663	5	4.7	173	9	US-10-187-751-506	Sequence 506, App
591	5	4.7	173	9	US-10-173-705-506	Sequence 506, App	664	5	4.7	173	9	US-10-187-753-506	Sequence 506, App
592	5	4.7	173	9	US-10-174-576-506	Sequence 506, App	665	5	4.7	173	9	US-10-187-754-506	Sequence 506, App
593	5	4.7	173	9	US-10-174-585-506	Sequence 506, App	666	5	4.7	173	9	US-10-187-757-506	Sequence 506, App
594	5	4.7	173	9	US-10-174-586-506	Sequence 506, App	667	5	4.7	173	9	US-10-187-884-506	Sequence 506, App
595	5	4.7	173	9	US-10-175-747-506	Sequence 506, App	668	5	4.7	173	9	US-10-188-767-506	Sequence 506, App
596	5	4.7	173	9	US-10-176-481-506	Sequence 506, App	669	5	4.7	173	9	US-10-188-769-506	Sequence 506, App
597	5	4.7	173	9	US-10-176-485-506	Sequence 506, App	670	5	4.7	173	9	US-10-188-770-506	Sequence 506, App
598	5	4.7	173	9	US-10-176-487-506	Sequence 506, App	671	5	4.7	173	9	US-10-188-773-506	Sequence 506, App
599	5	4.7	173	9	US-10-176-493-506	Sequence 506, App	672	5	4.7	173	9	US-10-188-781-506	Sequence 506, App
600	5	4.7	173	9	US-10-176-756-506	Sequence 506, App	673	5	4.7	173	9	US-10-194-361-506	Sequence 506, App
601	5	4.7	173	9	US-10-176-911-506	Sequence 506, App	674	5	4.7	173	9	US-10-194-423-506	Sequence 506, App
602	5	4.7	173	9	US-10-176-919-506	Sequence 506, App	675	5	4.7	173	9	US-10-195-897-506	Sequence 506, App
603	5	4.7	173	9	US-10-176-925-506	Sequence 506, App	676	5	4.7	173	9	US-10-195-901-506	Sequence 506, App

677	5	4.7	173	9	US-10-196-756-506	Sequence 506, App	750	5	4.7	173	9	US-10-199-310-506	Sequence 506, App
678	5	4.7	173	9	US-10-173-708-506	Sequence 506, App	751	5	4.7	173	9	US-10-199-311-506	Sequence 506, App
679	5	4.7	173	9	US-10-176-479-506	Sequence 506, App	752	5	4.7	173	9	US-10-199-314-506	Sequence 506, App
680	5	4.7	173	9	US-10-176-748-506	Sequence 506, App	753	5	4.7	173	9	US-10-199-317-506	Sequence 506, App
681	5	4.7	173	9	US-10-176-916-506	Sequence 506, App	754	5	4.7	173	9	US-10-199-665-506	Sequence 506, App
682	5	4.7	173	9	US-10-179-507-506	Sequence 506, App	755	5	4.7	173	9	US-10-199-666-506	Sequence 506, App
683	5	4.7	173	9	US-10-179-516-506	Sequence 506, App	756	5	4.7	173	9	US-10-199-669-506	Sequence 506, App
684	5	4.7	173	9	US-10-179-519-506	Sequence 506, App	757	5	4.7	173	9	US-10-201-534-506	Sequence 506, App
685	5	4.7	173	9	US-10-179-525-506	Sequence 506, App	758	5	4.7	173	9	US-10-201-770-506	Sequence 506, App
686	5	4.7	173	9	US-10-180-540-506	Sequence 506, App	759	5	4.7	173	9	US-10-201-855-506	Sequence 506, App
687	5	4.7	173	9	US-10-180-545-506	Sequence 506, App	760	5	4.7	173	9	US-10-201-856-506	Sequence 506, App
688	5	4.7	173	9	US-10-183-006-506	Sequence 506, App	761	5	4.7	173	9	US-10-202-469-506	Sequence 506, App
689	5	4.7	173	9	US-10-183-008-506	Sequence 506, App	762	5	4.7	173	9	US-10-202-470-506	Sequence 506, App
690	5	4.7	173	9	US-10-183-017-506	Sequence 506, App	763	5	4.7	173	9	US-10-202-476-506	Sequence 506, App
691	5	4.7	173	9	US-10-183-019-506	Sequence 506, App	764	5	4.7	173	9	US-10-202-934-506	Sequence 506, App
692	5	4.7	173	9	US-10-184-618-506	Sequence 506, App	765	5	4.7	173	9	US-10-202-935-506	Sequence 506, App
693	5	4.7	173	9	US-10-184-625-506	Sequence 506, App	766	5	4.7	173	9	US-10-202-936-506	Sequence 506, App
694	5	4.7	173	9	US-10-184-626-506	Sequence 506, App	767	5	4.7	173	9	US-10-202-939-506	Sequence 506, App
695	5	4.7	173	9	US-10-184-627-506	Sequence 506, App	768	5	4.7	173	9	US-10-205-504-506	Sequence 506, App
696	5	4.7	173	9	US-10-184-645-506	Sequence 506, App	769	5	4.7	173	9	US-10-205-509-506	Sequence 506, App
697	5	4.7	173	9	US-10-184-654-506	Sequence 506, App	770	5	4.7	173	9	US-10-205-895-506	Sequence 506, App
698	5	4.7	173	9	US-10-184-655-506	Sequence 506, App	771	5	4.7	173	9	US-10-205-899-506	Sequence 506, App
699	5	4.7	173	9	US-10-188-774-506	Sequence 506, App	772	5	4.7	173	9	US-10-205-900-506	Sequence 506, App
700	5	4.7	173	9	US-10-188-775-506	Sequence 506, App	773	5	4.7	173	9	US-10-205-909-506	Sequence 506, App
701	5	4.7	173	9	US-10-194-462-506	Sequence 506, App	774	5	4.7	173	9	US-10-183-002-506	Sequence 506, App
702	5	4.7	173	9	US-10-195-902-506	Sequence 506, App	775	5	4.7	173	9	US-10-184-631-506	Sequence 506, App
703	5	4.7	173	9	US-10-196-743-506	Sequence 506, App	776	5	4.7	173	9	US-10-184-638-506	Sequence 506, App
704	5	4.7	173	9	US-10-196-745-506	Sequence 506, App	777	5	4.7	173	9	US-10-187-752-506	Sequence 506, App
705	5	4.7	173	9	US-10-196-760-506	Sequence 506, App	778	5	4.7	173	9	US-10-187-787-506	Sequence 506, App
706	5	4.7	173	9	US-10-196-762-506	Sequence 506, App	779	5	4.7	173	9	US-10-194-461-506	Sequence 506, App
707	5	4.7	173	9	US-10-197-695-506	Sequence 506, App	780	5	4.7	173	9	US-10-195-882-506	Sequence 506, App
708	5	4.7	173	9	US-10-176-484-506	Sequence 506, App	781	5	4.7	173	9	US-10-196-751-506	Sequence 506, App
709	5	4.7	173	9	US-10-176-753-506	Sequence 506, App	782	5	4.7	173	9	US-10-197-694-506	Sequence 506, App
710	5	4.7	173	9	US-10-176-917-506	Sequence 506, App	783	5	4.7	173	9	US-10-197-697-506	Sequence 506, App
711	5	4.7	173	9	US-10-176-982-506	Sequence 506, App	784	5	4.7	173	9	US-10-197-707-506	Sequence 506, App
712	5	4.7	173	9	US-10-179-506-506	Sequence 506, App	785	5	4.7	173	9	US-10-199-303-506	Sequence 506, App
713	5	4.7	173	9	US-10-179-513-506	Sequence 506, App	786	5	4.7	173	9	US-10-199-318-506	Sequence 506, App
714	5	4.7	173	9	US-10-179-514-506	Sequence 506, App	787	5	4.7	173	9	US-10-199-482-506	Sequence 506, App
715	5	4.7	173	9	US-10-179-522-506	Sequence 506, App	788	5	4.7	173	9	US-10-199-482-506	Sequence 506, App
716	5	4.7	173	9	US-10-180-556-506	Sequence 506, App	789	5	4.7	173	9	US-10-201-324-506	Sequence 506, App
717	5	4.7	173	9	US-10-180-556-506	Sequence 506, App	790	5	4.7	173	9	US-10-201-328-506	Sequence 506, App
718	5	4.7	173	9	US-10-183-015-506	Sequence 506, App	791	5	4.7	173	9	US-10-201-527-506	Sequence 506, App
719	5	4.7	173	9	US-10-184-615-506	Sequence 506, App	792	5	4.7	173	9	US-10-201-528-506	Sequence 506, App
720	5	4.7	173	9	US-10-184-620-506	Sequence 506, App	793	5	4.7	173	9	US-10-201-529-506	Sequence 506, App
721	5	4.7	173	9	US-10-184-643-506	Sequence 506, App	794	5	4.7	173	9	US-10-201-530-506	Sequence 506, App
722	5	4.7	173	9	US-10-184-656-506	Sequence 506, App	795	5	4.7	173	9	US-10-202-408-506	Sequence 506, App
723	5	4.7	173	9	US-10-192-010-506	Sequence 506, App	796	5	4.7	173	9	US-10-202-409-506	Sequence 506, App
724	5	4.7	173	9	US-10-195-894-506	Sequence 506, App	797	5	4.7	173	9	US-10-202-411-506	Sequence 506, App
725	5	4.7	173	9	US-10-205-908-506	Sequence 506, App	798	5	4.7	173	9	US-10-202-472-506	Sequence 506, App
726	5	4.7	173	9	US-10-184-619-506	Sequence 506, App	799	5	4.7	173	9	US-10-205-507-506	Sequence 506, App
727	5	4.7	173	9	US-10-186-885-506	Sequence 506, App	800	5	4.7	173	9	US-10-205-511-506	Sequence 506, App
728	5	4.7	173	9	US-10-187-750-506	Sequence 506, App	801	5	4.7	173	9	US-10-205-922-506	Sequence 506, App
729	5	4.7	173	9	US-10-188-780-506	Sequence 506, App	802	5	4.7	173	9	US-10-205-907-506	Sequence 506, App
730	5	4.7	173	9	US-10-192-015-506	Sequence 506, App	803	5	4.7	173	9	US-10-194-456-506	Sequence 506, App
731	5	4.7	173	9	US-10-194-394-506	Sequence 506, App	804	5	4.7	173	9	US-10-196-758-506	Sequence 506, App
732	5	4.7	173	9	US-10-194-425-506	Sequence 506, App	805	5	4.7	173	9	US-10-198-770-506	Sequence 506, App
733	5	4.7	173	9	US-10-194-425-506	Sequence 506, App	806	5	4.7	173	9	US-10-198-770-506	Sequence 506, App
734	5	4.7	173	9	US-10-194-485-506	Sequence 506, App	807	5	4.7	173	9	US-10-199-308-506	Sequence 506, App
735	5	4.7	173	9	US-10-195-885-506	Sequence 506, App	808	5	4.7	173	9	US-10-200-617-506	Sequence 506, App
736	5	4.7	173	9	US-10-195-890-506	Sequence 506, App	809	5	4.7	173	9	US-10-205-893-506	Sequence 506, App
737	5	4.7	173	9	US-10-195-899-506	Sequence 506, App	810	5	4.7	173	9	US-10-205-897-506	Sequence 506, App
738	5	4.7	173	9	US-10-196-748-506	Sequence 506, App	811	5	4.7	173	9	US-09-915-582-97	Sequence 97, App1
739	5	4.7	173	9	US-10-196-750-506	Sequence 506, App	812	5	4.7	173	12	US-10-052-586-506	Sequence 506, App
740	5	4.7	173	9	US-10-197-699-506	Sequence 506, App	813	5	4.7	176	9	US-09-726-626-3525	Sequence 3525, App
741	5	4.7	173	9	US-10-197-700-506	Sequence 506, App	814	5	4.7	176	10	US-09-216-293-73	Sequence 73, App1
742	5	4.7	173	9	US-10-197-705-506	Sequence 506, App	815	5	4.7	176	10	US-09-788-200-6	Sequence 6, App11
743	5	4.7	173	9	US-10-197-708-506	Sequence 506, App	816	5	4.7	176	10	US-09-791-171-50	Sequence 50, App1
744	5	4.7	173	9	US-10-198-764-506	Sequence 506, App	817	5	4.7	177	10	US-09-734-0178-70	Sequence 70, App1
745	5	4.7	173	9	US-10-198-765-506	Sequence 506, App	818	5	4.7	181	9	US-10-081-051-37	Sequence 37, App1
746	5	4.7	173	9	US-10-198-766-506	Sequence 506, App	819	5	4.7	181	10	US-09-815-242-11589	Sequence 11589, A
747	5	4.7	173	9	US-10-198-769-506	Sequence 506, App	820	5	4.7	181	12	US-10-096-241-4	Sequence 4, App1i
748	5	4.7	173	9	US-10-199-305-506	Sequence 506, App	821	5	4.7	183	9	US-09-764-868-993	Sequence 893, App
749	5	4.7	173	9	US-10-199-306-506	Sequence 506, App	822	5	4.7	186	9	US-10-214-811-14	Sequence 14, App1



```
969 5 4.7 293 9 US-09-738-626-5474 Sequence 5474, Ap
970 5 4.7 293 10 US-09-968-958-2 Sequence 2, Appli
971 5 4.7 294 9 US-10-098-807-6 Sequence 6, Appli
972 5 4.7 294 10 US-09-788-626-27 Sequence 27, Appl
973 5 4.7 296 9 US-10-098-807-7 Sequence 7, Appli
974 5 4.7 297 9 US-10-098-807-5 Sequence 5, Appli
975 5 4.7 297 10 US-09-816-028A-31 Sequence 31, Appli
976 5 4.7 297 10 US-09-775-009-2 Sequence 2, Appli
977 5 4.7 299 10 US-09-861-451A-36 Sequence 36, Appli
978 5 4.7 300 9 US-10-098-807-3 Sequence 3, Appli
979 5 4.7 300 9 US-10-098-807-4 Sequence 4, Appli
980 5 4.7 300 10 US-09-925-301-1008 Sequence 1008, Ap
981 5 4.7 301 9 US-10-080-960-11 Sequence 11, Appli
982 5 4.7 301 9 US-10-116-255-22 Sequence 22, Appli
983 5 4.7 301 9 US-09-925-299-786 Sequence 786, Appli
984 5 4.7 301 10 US-09-816-028A-27 Sequence 27, Appli
985 5 4.7 301 10 US-09-925-299-786 Sequence 786, Appli
986 5 4.7 302 10 US-09-948-078-2 Sequence 2, Appli
987 5 4.7 303 10 US-09-925-302-674 Sequence 674, Appli
988 5 4.7 304 9 US-09-738-626-6082 Sequence 6082, Ap
989 5 4.7 305 9 US-10-162-012-23 Sequence 23, Appli
990 5 4.7 305 10 US-09-875-363-4 Sequence 4, Appli
991 5 4.7 305 10 US-09-771-730-119 Sequence 119, Appli
992 5 4.7 305 10 US-09-925-300-1194 Sequence 1194, Ap
993 5 4.7 308 9 US-09-809-391-607 Sequence 607, Appli
994 5 4.7 308 10 US-09-999-162-2 Sequence 2, Appli
995 5 4.7 309 9 US-09-510-332-147 Sequence 147, Appli
996 5 4.7 309 9 US-09-510-332-151 Sequence 151, Appli
997 5 4.7 309 10 US-09-788-626-16 Sequence 16, Appli
998 5 4.7 309 10 US-09-801-368-290 Sequence 290, Appli
999 5 4.7 310 10 US-09-815-242-12090 Sequence 12090, A
1000 5 4.7 312 9 US-10-084-205-14 Sequence 14, Appli
```

## ALIGNMENTS

```
RESULT 1
US-10-020-441-2
; Sequence 2, Application US/10020441
; Publication No. US20020182224A1
; GENERAL INFORMATION:
; APPLICANT: Doenhoff, Michael
; APPLICANT: Sayers, Jon
; TITLE OF INVENTION: SCHISTOSOMIASIS VACCINE
; FILE REFERENCE: Doenhoff
; CURRENT APPLICATION NUMBER: US/10/020,441
; CURRENT FILING DATE: 2001-12-18
; PRIOR APPLICATION NUMBER: US/09/413,810
; PRIOR FILING DATE: PRIOR FILING DATE: 1999-10-07
; NUMBER OF SEQ ID NOS: 3
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2
; LENGTH: 106
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: protein
US-10-020-441-2
```

```
Query Match 100.0%; Score 106; DB 9; Length 106;
Best Local Similarity 100.0%; Pred. No. 2.5e-99;
Matches 106; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 1 VSFILRNDDGQIHQSGVAVAPGYMPCMSARORRIIAOTLSGFDIAIWMAMVNL 60
DB 1 VSFILRNDDGQIHQSGVAVAPGYMPCMSARORRIIAOTLSGFDIAIWMAMVNL 60
QY 61 QSGIRVYISLPPSPDIPPPGTGVFIVGYGRDDNDPDRSRKNGGLKK 106
DB 61 QSGIRVYISLPPSPDIPPPGTGVFIVGYGRDDNDPDRSRKNGGLKK 106
```

```
RESULT 2
US-10-020-441-3
; Sequence 3, Application US/10020441
; Publication No. US20020182224A1
; GENERAL INFORMATION:
; APPLICANT: Doenhoff, Michael
; APPLICANT: Sayers, Jon
; TITLE OF INVENTION: SCHISTOSOMIASIS VACCINE
; FILE REFERENCE: Doenhoff
; CURRENT APPLICATION NUMBER: US/10/020,441
; CURRENT FILING DATE: 2001-12-18
; PRIOR APPLICATION NUMBER: US/09/413,810
; PRIOR FILING DATE: PRIOR FILING DATE: 1999-10-07
; NUMBER OF SEQ ID NOS: 3
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 3
; LENGTH: 16
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: protein
US-10-020-441-3
```

```
Query Match 15.1%; Score 16; DB 9; Length 16;
Best Local Similarity 100.0%; Pred. No. 1.5e-09;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 85 VGYGRDDNDPDRSRKN 100
DB 1 VGYGRDDNDPDRSRKN 16
```

```
RESULT 3
US-10-016-668-4
; Sequence 4, Application US/10016668
; Publication No. US20030032059A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Zhen-Gang
; APPLICANT: Voigt, Christopher A.
; APPLICANT: Mayo, Stephen L.
; APPLICANT: Arnold, Frances H.
; TITLE OF INVENTION: GENE RECOMBINATION AND HYBRID PROTEIN DEVELOPMENT
; FILE REFERENCE: 9373/1H812-US3
; CURRENT APPLICATION NUMBER: US/10/016,668
; CURRENT FILING DATE: 2002-04-22
; PRIOR APPLICATION NUMBER: US 09/863,765
; PRIOR FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: US 60/207,048
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: US 60/235,960
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: US 60/283,567
; PRIOR FILING DATE: 2001-04-13
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 4
; LENGTH: 359
; TYPE: PRT
; ORGANISM: Klebsiella pneumoniae
; PUBLICATION INFORMATION:
; DATABASE ACCESSION NUMBER: SWISPROT / Q48437
; DATABASE ENTRY DATE: 1996-11-01
; RELEVANT RESIDUES: (1)...(359)
US-10-016-668-4
```

```
Query Match 6.6%; Score 7; DB 9; Length 359;
Best Local Similarity 100.0%; Pred. No. 27;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 49 IAIWMLA 55
DB 332 IAIWMLA 338
```

RESULT 4  
US-09-808-602-83  
; Sequence 83, Application US/09808602  
; Patent No. US20020155115A1  
; GENERAL INFORMATION:  
; APPLICANT: Vernet, Corine A  
; APPLICANT: Fernandes, Elma  
; APPLICANT: Shimkets, Richard A  
; APPLICANT: Herrman, John L  
; APPLICANT: Majumder, Kumud  
; APPLICANT: Mishra, Vishnu  
; APPLICANT: Mezes, Peter S  
; APPLICANT: MacDougall, John  
; TITLE OF INVENTION: No. US20020155115A1 Proteins and Nucleic Acids Encoding Same  
; FILE REFERENCE: 15966-697 CIP  
; CURRENT APPLICATION NUMBER: US/09/808,602  
; CURRENT FILING DATE: 2001-03-14  
; PRIOR APPLICATION NUMBER: 09/800,198  
; PRIOR FILING DATE: 2001-03-05  
; PRIOR APPLICATION NUMBER: 60/186,596  
; PRIOR FILING DATE: 2000-03-03  
; NUMBER OF SEQ ID NOS: 114  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 83  
; LENGTH: 1737  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-808-602-83

Query Match 6.6%; Score 7; DB 9; Length 1737;  
Best Local Similarity 100.0%; Pred. No. 1.1e+02;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 57 MVNLOSG 63  
|||||  
Db 1014 MVNLOSG 1020

RESULT 5  
US-09-808-602-13  
; Sequence 13, Application US/09808602  
; Patent No. US20020155115A1  
; GENERAL INFORMATION:  
; APPLICANT: Vernet, Corine A  
; APPLICANT: Fernandes, Elma  
; APPLICANT: Shimkets, Richard A  
; APPLICANT: Herrman, John L  
; APPLICANT: Majumder, Kumud  
; APPLICANT: Mishra, Vishnu  
; APPLICANT: Mezes, Peter S  
; APPLICANT: MacDougall, John  
; TITLE OF INVENTION: No. US20020155115A1 Proteins and Nucleic Acids Encoding Same  
; FILE REFERENCE: 15966-697 CIP  
; CURRENT APPLICATION NUMBER: US/09/808,602  
; CURRENT FILING DATE: 2001-03-14  
; PRIOR APPLICATION NUMBER: 09/800,198  
; PRIOR FILING DATE: 2001-03-05  
; PRIOR APPLICATION NUMBER: 60/186,596  
; PRIOR FILING DATE: 2000-03-03  
; NUMBER OF SEQ ID NOS: 114  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 13  
; LENGTH: 2724  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-808-602-13

Query Match 6.6%; Score 7; DB 9; Length 2724;  
Best Local Similarity 100.0%; Pred. No. 1.6e+02;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 57 MVNLOSG 63

Db 2001 MVNLOSG 2007  
|||||  
RESULT 6  
US-09-808-602-8  
; Sequence 8, Application US/09808602  
; Patent No. US20020155115A1  
; GENERAL INFORMATION:  
; APPLICANT: Vernet, Corine A  
; APPLICANT: Fernandes, Elma  
; APPLICANT: Shimkets, Richard A  
; APPLICANT: Herrman, John L  
; APPLICANT: Majumder, Kumud  
; APPLICANT: Mishra, Vishnu  
; APPLICANT: Mezes, Peter S  
; APPLICANT: MacDougall, John  
; TITLE OF INVENTION: No. US20020155115A1 Proteins and Nucleic Acids Encoding Same  
; FILE REFERENCE: 15966-697 CIP  
; CURRENT APPLICATION NUMBER: US/09/808,602  
; CURRENT FILING DATE: 2001-03-14  
; PRIOR APPLICATION NUMBER: 09/800,198  
; PRIOR FILING DATE: 2001-03-05  
; PRIOR APPLICATION NUMBER: 60/186,596  
; PRIOR FILING DATE: 2000-03-03  
; NUMBER OF SEQ ID NOS: 114  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 8  
; LENGTH: 2733  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-808-602-8

Query Match 6.6%; Score 7; DB 9; Length 2733;  
Best Local Similarity 100.0%; Pred. No. 1.6e+02;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 57 MVNLOSG 63  
|||||  
Db 2010 MVNLOSG 2016

RESULT 7  
US-09-808-602-80  
; Sequence 80, Application US/09808602  
; Patent No. US20020155115A1  
; GENERAL INFORMATION:  
; APPLICANT: Vernet, Corine A  
; APPLICANT: Fernandes, Elma  
; APPLICANT: Shimkets, Richard A  
; APPLICANT: Herrman, John L  
; APPLICANT: Majumder, Kumud  
; APPLICANT: Mishra, Vishnu  
; APPLICANT: Mezes, Peter S  
; APPLICANT: MacDougall, John  
; TITLE OF INVENTION: No. US20020155115A1 Proteins and Nucleic Acids Encoding Same  
; FILE REFERENCE: 15966-697 CIP  
; CURRENT APPLICATION NUMBER: US/09/808,602  
; CURRENT FILING DATE: 2001-03-14  
; PRIOR APPLICATION NUMBER: 09/800,198  
; PRIOR FILING DATE: 2001-03-05  
; PRIOR APPLICATION NUMBER: 60/186,596  
; PRIOR FILING DATE: 2000-03-03  
; NUMBER OF SEQ ID NOS: 114  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 80  
; LENGTH: 2764  
; TYPE: PRT  
; ORGANISM: Drosophila melanogaster  
US-09-808-602-80

Query Match 6.6%; Score 7; DB 9; Length 2764;  
Best Local Similarity 100.0%; Pred. No. 1.7e+02;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 57 MVNLOSG 63  
|||||

Db 2041 MVNLOSG 2047

## RESULT 8

US-09-808-602-84  
; Sequence 84, Application US/09808602  
; Patent No. US20020155115A1  
; GENERAL INFORMATION:  
; APPLICANT: Vernet, Corine A  
; APPLICANT: Fernandes, Elma  
; APPLICANT: Shimkets, Richard A  
; APPLICANT: Herrman, John L  
; APPLICANT: Majumder, Kunud  
; APPLICANT: Mishra, Vishnu  
; APPLICANT: Mezes, Peter S  
; APPLICANT: MacDougall, John  
; TITLE OF INVENTION: No. US20020155115A1el Proteins and Nucleic Acids Encoding Same  
; FILE REFERENCE: 15966-697 CIP  
; CURRENT APPLICATION NUMBER: US/09/808,602  
; CURRENT FILING DATE: 2001-03-14  
; PRIOR APPLICATION NUMBER: 09/800,198  
; PRIOR FILING DATE: 2001-03-05  
; PRIOR APPLICATION NUMBER: 60/186,596  
; PRIOR FILING DATE: 2000-03-03  
; NUMBER OF SEQ ID NOS: 114  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 84  
; LENGTH: 2765  
; TYPE: PRT  
; ORGANISM: Rattus norvegicus  
US-09-808-602-84

Query Match 6.6%; Score 7; DB 9; Length 2765;  
Best Local Similarity 100.0%; Pred. No. 1.7e+02;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 57 MVNLOSG 63  
|||||

Db 2042 MVNLOSG 2048

## RESULT 9

US-09-808-602-81  
; Sequence 81, Application US/09808602  
; Patent No. US20020155115A1  
; GENERAL INFORMATION:  
; APPLICANT: Vernet, Corine A  
; APPLICANT: Fernandes, Elma  
; APPLICANT: Shimkets, Richard A  
; APPLICANT: Herrman, John L  
; APPLICANT: Majumder, Kunud  
; APPLICANT: Mishra, Vishnu  
; APPLICANT: Mezes, Peter S  
; APPLICANT: MacDougall, John  
; TITLE OF INVENTION: No. US20020155115A1el Proteins and Nucleic Acids Encoding Same  
; FILE REFERENCE: 15966-697 CIP  
; CURRENT APPLICATION NUMBER: US/09/808,602  
; CURRENT FILING DATE: 2001-03-14  
; PRIOR APPLICATION NUMBER: 09/800,198  
; PRIOR FILING DATE: 2001-03-05  
; PRIOR APPLICATION NUMBER: 60/186,596  
; PRIOR FILING DATE: 2000-03-03  
; NUMBER OF SEQ ID NOS: 114  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 81  
; LENGTH: 2802  
; TYPE: PRT  
; ORGANISM: Gallus gallus  
US-09-808-602-81

Query Match 6.6%; Score 7; DB 9; Length 2802;  
Best Local Similarity 100.0%; Pred. No. 1.7e+02;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 57 MVNLOSG 63  
|||||

Db 2079 MVNLOSG 2085

## RESULT 10

US-09-027-956-6  
; Sequence 6, Application US/09027956A  
; Patent No. US20010048929A1  
; GENERAL INFORMATION:  
; APPLICANT: CHONG, Pele  
; APPLICANT: LINDBERG, Alf  
; APPLICANT: KLIN, Michel  
; TITLE OF INVENTION: NOVEL MULTI-OLIGOSACCHARIDE GLYCOCONJUGATE BACTERIAL  
; FILE REFERENCE: 1038-791 MIS:jb  
; CURRENT APPLICATION NUMBER: US/09/027,956A  
; CURRENT FILING DATE: 1998-02-23  
; NUMBER OF SEQ ID NOS: 8  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 6  
; LENGTH: 21  
; TYPE: PRT  
; ORGANISM: Streptococcus pneumoniae  
US-09-027-956-6

Query Match 5.7%; Score 6; DB 10; Length 21;  
Best Local Similarity 100.0%; Pred. No. 21;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 20 GVKVAP 25  
|||||

Db 8 GVKVAP 13

RESULT 11  
US-09-864-761-47769  
; Sequence 47769, Application US/09864761  
; Patent No. US20020048763A1  
; GENERAL INFORMATION:  
; APPLICANT: Penn, Sharon G.  
; APPLICANT: Rank, David R.  
; APPLICANT: Hanzel, David K.  
; APPLICANT: Chen, Wensheng  
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR  
; FILE REFERENCE: Aomicca-X-1  
; CURRENT APPLICATION NUMBER: US/09/864,761  
; CURRENT FILING DATE: 2001-05-23  
; PRIOR APPLICATION NUMBER: US 60/180,312  
; PRIOR FILING DATE: 2000-02-04  
; PRIOR APPLICATION NUMBER: US 60/207,456  
; PRIOR FILING DATE: 2000-05-26  
; PRIOR APPLICATION NUMBER: US 09/632,366  
; PRIOR FILING DATE: 2000-08-03  
; PRIOR APPLICATION NUMBER: GB 24263,6  
; PRIOR FILING DATE: 2000-10-04  
; PRIOR APPLICATION NUMBER: US 60/236,359  
; PRIOR FILING DATE: 2000-09-27  
; PRIOR APPLICATION NUMBER: PCT/US01/00666  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00667  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00664  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00669  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00665



```
;; PRIOR FILING DATE: 2001-01-30
;; PRIOR APPLICATION NUMBER: PCT/US01/00668
;; PRIOR FILING DATE: 2001-01-30
;; PRIOR APPLICATION NUMBER: PCT/US01/00663
;; PRIOR FILING DATE: 2001-01-30
;; PRIOR APPLICATION NUMBER: PCT/US01/00662
;; PRIOR FILING DATE: 2001-01-30
;; PRIOR APPLICATION NUMBER: PCT/US01/00661
;; PRIOR FILING DATE: 2001-01-30
;; PRIOR APPLICATION NUMBER: PCT/US01/00670
;; PRIOR FILING DATE: 2001-01-30
;; PRIOR APPLICATION NUMBER: US 60/234,687
;; PRIOR FILING DATE: 2000-09-21
;; PRIOR APPLICATION NUMBER: US 09/608,408
;; PRIOR FILING DATE: 2000-06-30
;; PRIOR APPLICATION NUMBER: US 09/774,203
;; PRIOR FILING DATE: 2001-01-29
;; NUMBER OF SEQ ID NOS: 49117
;; SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
;; SEQ ID NO 47769
;; LENGTH: 30
;; TYPE: PRT
;; ORGANISM: Homo sapiens
;; FEATURE:
;; OTHER INFORMATION: MAP TO AC011619.1
;; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 13
;; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 0.98
US-09-864-761-47769
```

```
Query Match      5.7%; Score 6; DB 10; Length 30;
Best Local Similarity 100.0%; Pred. No. 29;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
Qy      92 NDRDPS 97
Db      6 NDRDPS 11
```

```
RESULT 12
US-09-864-761-38283
;; Sequence 38283, Application US/09864761
;; Patent No. US20020048763A1
;; GENERAL INFORMATION:
;; APPLICANT: Penn, Sharon G.
;; APPLICANT: Rank, David R.
;; APPLICANT: Hanzel, David K.
;; APPLICANT: Chen, Wensheng
;; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
;; FILE REFERENCE: Aecm1ca-X-1
;; CURRENT APPLICATION NUMBER: US/09/864,761
;; CURRENT FILING DATE: 2001-05-23
;; PRIOR APPLICATION NUMBER: US 60/180,312
;; PRIOR FILING DATE: 2000-02-04
;; PRIOR APPLICATION NUMBER: US 60/207,456
;; PRIOR FILING DATE: 2000-05-26
;; PRIOR APPLICATION NUMBER: US 09/632,366
;; PRIOR FILING DATE: 2000-08-03
;; PRIOR APPLICATION NUMBER: GB 24263, 6
;; PRIOR FILING DATE: 2000-10-04
;; PRIOR APPLICATION NUMBER: US 60/236,359
;; PRIOR FILING DATE: 2000-09-27
;; PRIOR APPLICATION NUMBER: PCT/US01/00666
;; PRIOR FILING DATE: 2001-01-30
;; PRIOR APPLICATION NUMBER: PCT/US01/00667
;; PRIOR FILING DATE: 2001-01-30
;; PRIOR APPLICATION NUMBER: PCT/US01/00664
;; PRIOR FILING DATE: 2001-01-30
;; PRIOR APPLICATION NUMBER: PCT/US01/00669
;; PRIOR FILING DATE: 2001-01-30
;; PRIOR APPLICATION NUMBER: PCT/US01/00665
;; PRIOR FILING DATE: 2001-01-30
;; PRIOR APPLICATION NUMBER: PCT/US01/00668
```

```
;; PRIOR FILING DATE: 2001-01-30
;; PRIOR APPLICATION NUMBER: PCT/US01/00663
;; PRIOR FILING DATE: 2001-01-30
;; PRIOR APPLICATION NUMBER: PCT/US01/00662
;; PRIOR FILING DATE: 2001-01-30
;; PRIOR APPLICATION NUMBER: PCT/US01/00661
;; PRIOR FILING DATE: 2001-01-30
;; PRIOR APPLICATION NUMBER: PCT/US01/00670
;; PRIOR FILING DATE: 2001-01-30
;; PRIOR APPLICATION NUMBER: US 60/234,687
;; PRIOR FILING DATE: 2000-09-21
;; PRIOR APPLICATION NUMBER: US 09/608,408
;; PRIOR FILING DATE: 2000-06-30
;; PRIOR APPLICATION NUMBER: US 09/774,203
;; PRIOR FILING DATE: 2001-01-29
;; NUMBER OF SEQ ID NOS: 49117
;; SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
;; SEQ ID NO 38283
;; LENGTH: 34
;; TYPE: PRT
;; ORGANISM: Homo sapiens
;; FEATURE:
;; OTHER INFORMATION: MAP TO AC012109.1
;; OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 0.99
US-09-864-761-38283
```

```
Query Match      5.7%; Score 6; DB 10; Length 34;
Best Local Similarity 100.0%; Pred. No. 33;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
Qy      61 QSGIRV 66
Db      8 QSGIRV 13
```

```
RESULT 13
US-09-796-692-1791
;; Sequence 1791, Application US/09796692
;; Publication No. US20020198362A1
;; GENERAL INFORMATION:
;; APPLICANT: Gaiger, Alexander
;; APPLICANT: Algate, Paul A.
;; APPLICANT: Mannion, Jane
;; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE DETECTION, DIAGNOSIS AND THERAP
;; FILE REFERENCE: 2077, 001200
;; CURRENT APPLICATION NUMBER: US/09/796,692
;; CURRENT FILING DATE: 2001-03-01
;; PRIOR APPLICATION NUMBER: 60/186,126
;; PRIOR FILING DATE: 2000-03-01
;; PRIOR APPLICATION NUMBER: 60/190,479
;; PRIOR FILING DATE: 2000-03-17
;; PRIOR APPLICATION NUMBER: 60/200,545
;; PRIOR FILING DATE: 2000-04-27
;; PRIOR APPLICATION NUMBER: 60/200,303
;; PRIOR FILING DATE: 2000-04-28
;; PRIOR APPLICATION NUMBER: 60/200,779
;; PRIOR FILING DATE: 2000-04-28
;; PRIOR APPLICATION NUMBER: 60/200,999
;; PRIOR FILING DATE: 2000-05-01
;; PRIOR APPLICATION NUMBER: 60/202,084
;; PRIOR FILING DATE: 2000-05-04
;; PRIOR APPLICATION NUMBER: 60/206,201
;; PRIOR FILING DATE: 2000-05-22
;; PRIOR APPLICATION NUMBER: 60/218,950
;; PRIOR FILING DATE: 2000-07-14
;; PRIOR APPLICATION NUMBER: 60/222,903
;; PRIOR FILING DATE: 2000-08-03
;; PRIOR APPLICATION NUMBER: 60/223,416
;; PRIOR FILING DATE: 2000-08-04
;; PRIOR APPLICATION NUMBER: 60/223,378
;; PRIOR FILING DATE: 2000-08-07
;; NUMBER OF SEQ ID NOS: 9597
```

```
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 1791
; LENGTH: 75
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-796-692-1791

Query Match
Best Local Similarity 100.0%; Score 6; DB 9; Length 75;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 98 RKNCGI 103
   |||||
Db 58 RKNCGI 63

RESULT 14
US-09-796-692-2055
; Sequence 2055, Application US/09796692
; Publication No. US20020198362A1
; GENERAL INFORMATION:
; APPLICANT: Gaiger, Alexander
; APPLICANT: Algate, Paul A.
; APPLICANT: Mannion, Jane
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE DETECTION, DIAGNOSIS AND THERAPY
; FILE REFERENCE: 2077,001200
; CURRENT APPLICATION NUMBER: US/09/796,692
; PRIOR FILING DATE: 2001-03-01
; PRIOR APPLICATION NUMBER: 60/186,126
; PRIOR FILING DATE: 2000-03-01
; PRIOR APPLICATION NUMBER: 60/190,479
; PRIOR FILING DATE: 2000-03-17
; PRIOR APPLICATION NUMBER: 60/200,545
; PRIOR FILING DATE: 2000-04-27
; PRIOR APPLICATION NUMBER: 60/200,303
; PRIOR FILING DATE: 2000-04-28
; PRIOR APPLICATION NUMBER: 60/200,779
; PRIOR FILING DATE: 2000-04-28
; PRIOR APPLICATION NUMBER: 60/200,999
; PRIOR FILING DATE: 2000-05-01
; PRIOR APPLICATION NUMBER: 60/202,084
; PRIOR FILING DATE: 2000-05-04
; PRIOR APPLICATION NUMBER: 60/206,201
; PRIOR FILING DATE: 2000-05-22
; PRIOR APPLICATION NUMBER: 60/218,950
; PRIOR FILING DATE: 2000-07-14
; PRIOR APPLICATION NUMBER: 60/222,903
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: 60/223,416
; PRIOR FILING DATE: 2000-08-04
; PRIOR APPLICATION NUMBER: 60/223,378
; PRIOR FILING DATE: 2000-08-07
; NUMBER OF SEQ ID NOS: 9597
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 2055
; LENGTH: 75
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-796-692-2055

Query Match
Best Local Similarity 5.7%; Score 6; DB 9; Length 75;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 98 RKNCGI 103
   |||||
Db 58 RKNCGI 63

RESULT 15
US-09-925-302-521
; Sequence 521, Application US/09925302

; Patent No. US20020044941A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
; FILE REFERENCE: PA104
; CURRENT APPLICATION NUMBER: US/09/925,302
; PRIOR FILING DATE: 2001-08-10
; PRIOR APPLICATION NUMBER: PCT/US00/05918
; PRIOR FILING DATE: 2000-03-08
; PRIOR APPLICATION NUMBER: 60/124,270
; PRIOR FILING DATE: 1999-03-12
; NUMBER OF SEQ ID NOS: 896
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 521
; LENGTH: 119
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (105)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-09-925-302-521

Query Match
Best Local Similarity 5.7%; Score 6; DB 10; Length 119;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 68 SLPOPS 73
   |||||
Db 32 SLPOPS 37

RESULT 16
US-09-864-761-36308
; Sequence 36308, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharon G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
; FILE REFERENCE: Aeo mica-X-1
; CURRENT APPLICATION NUMBER: US/09/864,761
; PRIOR FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/632,366
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: GB 24263,6
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US01/00666
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00664
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00669
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00665
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00668
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00663
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00662
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00661
```

```
;; PRIOR FILING DATE: 2001-01-30
;; PRIOR APPLICATION NUMBER: PCT/US01/00670
;; PRIOR FILING DATE: 2001-01-30
;; PRIOR APPLICATION NUMBER: US 60/234,687
;; PRIOR FILING DATE: 2000-09-21
;; PRIOR APPLICATION NUMBER: US 09/608,408
;; PRIOR FILING DATE: 2000-06-30
;; PRIOR APPLICATION NUMBER: US 09/774,203
;; PRIOR FILING DATE: 2001-01-29
;; NUMBER OF SEQ ID NOS: 49117
;; SOFTWARE: Annonax Sequence Listing Engine vers. 1.1
;; SEQ ID NO 36308
;; LENGTH: 128
;; TYPE: PRT
;; ORGANISM: Homo sapiens
;; FEATURE:
;; OTHER INFORMATION: MAP TO AL035669.28
;; OTHER INFORMATION: EXPRESSED IN HBL100, SIGNAL = 1
;; OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 1.2
;; OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 1.1
;; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 1.2
;; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 1
;; OTHER INFORMATION: EST HUMAN HIT: BF055116.1, EVALU2 2.00e-03
US-09-864-761-36308
```

```
Query Match          5.7%; Score 6; DB 10; Length 128;
Best Local Similarity 100.0%; Pred. No. 1.1e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY      68 SLP0PS 73
      |||||
Db      95 SLP0PS 100
```

```
RESULT 17
US-09-738-626-5580
;; Sequence 5580, Application US/09738626
;; Publication No. US20020197605A1
;; GENERAL INFORMATION:
;; APPLICANT: NAKAGAWA, SATOSHI
;; APPLICANT: MIZOGUCHI, HIROSHI
;; APPLICANT: ANDO, SEIKO
;; APPLICANT: HAYASHI, MIKIRO
;; APPLICANT: OCHIAI, KEIKO
;; APPLICANT: YOKOI, HARUHIKO
;; APPLICANT: TATEISHI, NAOKO
;; APPLICANT: SENOH, AKIHIRO
;; APPLICANT: IKEDA, MASATO
;; APPLICANT: OZAKI, AKIO
;; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
;; FILE REFERENCE: 249-125
;; CURRENT APPLICATION NUMBER: US/09/738, 626
;; PRIOR FILING DATE: 2000-12-18
;; PRIOR APPLICATION NUMBER: JP 99/377484
;; PRIOR FILING DATE: 1999-12-16
;; PRIOR APPLICATION NUMBER: JP 00/159162
;; PRIOR FILING DATE: 2000-04-07
;; PRIOR APPLICATION NUMBER: JP 00/280988
;; PRIOR FILING DATE: 2000-08-03
;; NUMBER OF SEQ ID NOS: 7059
;; SOFTWARE: PatentIn ver. 3.0
;; SEQ ID NO 5580
;; LENGTH: 144
;; TYPE: PRT
;; ORGANISM: Corynebacterium glutamicum
US-09-738-626-5580
```

```
Query Match          5.7%; Score 6; DB 9; Length 144;
Best Local Similarity 100.0%; Pred. No. 1.2e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY      41 AQTLSG 46
```

```
Db      67 AQTLSG 72
      |||||
```

```
RESULT 18
US-09-864-761-34045
;; Sequence 34045, Application US/09864761
;; Patent No. US20020048763A1
;; GENERAL INFORMATION:
;; APPLICANT: Penn, Sharon G.
;; APPLICANT: Rank, David R.
;; APPLICANT: Hanzel, David K.
;; APPLICANT: Chen, Wensheng
;; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
;; FILE REFERENCE: Aecmca-x-1
;; CURRENT APPLICATION NUMBER: US/09/864, 761
;; PRIOR FILING DATE: 2001-05-23
;; PRIOR APPLICATION NUMBER: US 60/180,312
;; PRIOR FILING DATE: 2000-02-04
;; PRIOR APPLICATION NUMBER: US 60/207,456
;; PRIOR FILING DATE: 2000-05-26
;; PRIOR APPLICATION NUMBER: US 09/632,366
;; PRIOR FILING DATE: 2000-08-03
;; PRIOR APPLICATION NUMBER: GB 24263, 6
;; PRIOR FILING DATE: 2000-10-04
;; PRIOR APPLICATION NUMBER: US 60/236,359
;; PRIOR FILING DATE: 2000-09-27
;; PRIOR APPLICATION NUMBER: PCT/US01/00666
;; PRIOR FILING DATE: 2001-01-30
;; PRIOR APPLICATION NUMBER: PCT/US01/00667
;; PRIOR FILING DATE: 2001-01-30
;; PRIOR APPLICATION NUMBER: PCT/US01/00664
;; PRIOR FILING DATE: 2001-01-30
;; PRIOR APPLICATION NUMBER: PCT/US01/00669
;; PRIOR FILING DATE: 2001-01-30
;; PRIOR APPLICATION NUMBER: PCT/US01/00665
;; PRIOR FILING DATE: 2001-01-30
;; PRIOR APPLICATION NUMBER: PCT/US01/00668
;; PRIOR FILING DATE: 2001-01-30
;; PRIOR APPLICATION NUMBER: PCT/US01/00663
;; PRIOR FILING DATE: 2001-01-30
;; PRIOR APPLICATION NUMBER: PCT/US01/00662
;; PRIOR FILING DATE: 2001-01-30
;; PRIOR APPLICATION NUMBER: PCT/US01/00661
;; PRIOR FILING DATE: 2001-01-30
;; PRIOR APPLICATION NUMBER: PCT/US01/00670
;; PRIOR FILING DATE: 2001-01-30
;; PRIOR APPLICATION NUMBER: US 60/234,687
;; PRIOR FILING DATE: 2000-09-21
;; PRIOR APPLICATION NUMBER: US 09/608,408
;; PRIOR FILING DATE: 2000-06-30
;; PRIOR APPLICATION NUMBER: US 09/774,203
;; PRIOR FILING DATE: 2001-01-29
;; NUMBER OF SEQ ID NOS: 49117
;; SOFTWARE: Annonax Sequence Listing Engine vers. 1.1
;; SEQ ID NO 34045
;; LENGTH: 146
;; TYPE: PRT
;; ORGANISM: Homo sapiens
;; FEATURE:
;; OTHER INFORMATION: MAP TO AC004967.2
;; OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 2.6
;; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 1.2
;; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 2.2
;; OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 1.7
;; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 1.5
;; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 1.5
;; OTHER INFORMATION: EXPRESSED IN HBL100, SIGNAL = 2
;; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 2.1
;; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 1.5
;; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 1.7
;; OTHER INFORMATION: EST HUMAN HIT: BF345274.1, EVALU2 5.00e-28
```

OTHER INFORMATION: EST HUMAN HIT: AA042813.1, EVALUATE 2.00e-26  
OTHER INFORMATION: SWISSPROT HIT: P35894, EVALUATE 1.00e-26  
US-09-864-761-34045

Query Match 5.7%; Score 6; DB 10; Length 146;  
Best Local Similarity 100.0%; Pred. No. 1.2e+02;

Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 75 IPPECT 80  
Db 2 IPPECT 7

RESULT 19  
US-09-975-719-158  
Sequence 158, Application US/09975719  
Publication No. US20030022349A1  
GENERAL INFORMATION:  
APPLICANT: Ausubel, Frederick M.  
APPLICANT: Rahme, Laurence G.  
TITLE OF INVENTION: VIRULENCE-ASSOCIATED NUCLEIC ACID  
FILE REFERENCE: 00786/361003  
CURRENT APPLICATION NUMBER: US/09/975,719  
CURRENT FILING DATE: 2001-10-10  
PRIOR APPLICATION NUMBER: US 09/199,637  
PRIOR FILING DATE: 1998-11-25  
PRIOR APPLICATION NUMBER: US 60/066,517  
PRIOR FILING DATE: 1997-11-25  
NUMBER OF SEQ ID NOS: 437  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 158  
LENGTH: 170  
TYPE: PRT  
ORGANISM: Pseudomonas aeruginosa  
US-09-975-719-158

Query Match 5.7%; Score 6; DB 9; Length 170;  
Best Local Similarity 100.0%; Pred. No. 1.4e+02;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 34 ARORRP 39  
Db 65 ARORRP 70

RESULT 20  
US-09-815-242-5164  
Sequence 5164, Application US/09815242  
Patent No. US20020061569A1  
GENERAL INFORMATION:  
APPLICANT: Haselbeck, Robert  
APPLICANT: Ohlsen, Kari L.  
APPLICANT: Zyckind, Judith W.  
APPLICANT: Wall, Daniel  
APPLICANT: Tirawick, John D.  
APPLICANT: Carr, Grant J.  
APPLICANT: Yamamoto, Robert T.  
APPLICANT: Xu, H. Howard  
TITLE OF INVENTION: Identification of Essential Genes in  
TITLE OF INVENTION: Prokaryotes  
FILE REFERENCE: ELITRA.01A  
CURRENT APPLICATION NUMBER: US/09/815,242  
CURRENT FILING DATE: 2001-03-21  
PRIOR APPLICATION NUMBER: 60/191,078  
PRIOR FILING DATE: 2000-03-21  
PRIOR APPLICATION NUMBER: 60/206,848  
PRIOR FILING DATE: 2000-05-23  
PRIOR APPLICATION NUMBER: 60/207,727  
PRIOR FILING DATE: 2000-05-26  
PRIOR APPLICATION NUMBER: 60/242,578  
PRIOR FILING DATE: 2000-10-23  
PRIOR APPLICATION NUMBER: 60/253,625

PRIOR FILING DATE: 2000-11-27  
PRIOR APPLICATION NUMBER: 60/257,931  
PRIOR FILING DATE: 2000-12-22  
PRIOR APPLICATION NUMBER: 60/269,308  
PRIOR FILING DATE: 2001-02-16  
NUMBER OF SEQ ID NOS: 14110  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 5164  
LENGTH: 177  
TYPE: PRT  
ORGANISM: Pseudomonas aeruginosa  
US-09-815-242-5164

Query Match 5.7%; Score 6; DB 10; Length 177;  
Best Local Similarity 100.0%; Pred. No. 1.4e+02;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 7 RRGDDQ 12  
Db 55 RRGDDQ 60

RESULT 21  
US-09-811-284-207  
Sequence 207, Application US/09811284  
Patent No. US20020058306A1  
GENERAL INFORMATION:  
APPLICANT: Vogel, Gabriel  
TITLE OF INVENTION: No. US20020058306A1 G Protein-Coupled Receptors  
FILE REFERENCE: 00167US1  
CURRENT APPLICATION NUMBER: US/09/811,284  
CURRENT FILING DATE: 2001-03-16  
PRIOR APPLICATION NUMBER: 60/189,783  
PRIOR FILING DATE: 2000-03-16  
PRIOR APPLICATION NUMBER: 60/189,907  
PRIOR FILING DATE: 2000-03-16  
PRIOR APPLICATION NUMBER: 60/189,918  
PRIOR FILING DATE: 2000-03-16  
PRIOR APPLICATION NUMBER: 60/189,960  
PRIOR FILING DATE: 2000-03-16  
PRIOR APPLICATION NUMBER: 60/189,917  
PRIOR FILING DATE: 2000-03-16  
PRIOR APPLICATION NUMBER: 60/192,945  
PRIOR FILING DATE: 2000-03-29  
PRIOR APPLICATION NUMBER: 60/192,916  
PRIOR FILING DATE: 2000-03-29  
PRIOR APPLICATION NUMBER: 60/192,923  
PRIOR FILING DATE: 2000-03-29  
PRIOR APPLICATION NUMBER: 60/192,933  
PRIOR FILING DATE: 2000-03-29  
PRIOR APPLICATION NUMBER: 60/192,830  
PRIOR FILING DATE: 2000-03-29  
PRIOR APPLICATION NUMBER: 60/192,234  
PRIOR FILING DATE: 2000-03-27  
PRIOR APPLICATION NUMBER: 60/192,155  
PRIOR FILING DATE: 2000-03-24  
PRIOR APPLICATION NUMBER: 60/192,935  
PRIOR FILING DATE: 2000-03-29  
NUMBER OF SEQ ID NOS: 258  
SOFTWARE: PatentIn version 3.0  
SEQ ID NO 207  
LENGTH: 196  
TYPE: PRT  
ORGANISM: Homo sapiens  
US-09-811-284-207

Query Match 5.7%; Score 6; DB 10; Length 196;  
Best Local Similarity 100.0%; Pred. No. 1.6e+02;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 82 VFIYGY 87  
Db 26 VFIYGY 31

## RESULT 22

```
US-09-738-626-3636
; Sequence 3636, Application US/097386626
; Publication No. US20020197605A1
; GENERAL INFORMATION:
; APPLICANT: NAKAGAWA, SATOSHI
; APPLICANT: MIZOGUCHI, HIROSHI
; APPLICANT: ANDO, SEIKO
; APPLICANT: HAYASHI, MIKIRO
; APPLICANT: OCHIAI, KEIKO
; APPLICANT: YOKOI, HARUHIKO
; APPLICANT: TATEISHI, NAOKO
; APPLICANT: SENOH, AKIHIRO
; APPLICANT: IKEDA, MASATO
; APPLICANT: OZAKI, AKIO
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-125
; CURRENT APPLICATION NUMBER: US/09/738,626
; CURRENT FILING DATE: 2000-12-18
; PRIOR APPLICATION NUMBER: JP 99/377484
; PRIOR FILING DATE: 1999-12-16
; PRIOR APPLICATION NUMBER: JP 00/159162
; PRIOR FILING DATE: 2000-04-07
; PRIOR APPLICATION NUMBER: JP 00/280988
; PRIOR FILING DATE: 2000-08-03
; NUMBER OF SEQ ID NOS: 7059
; SOFTWARE: PatentIn ver. 3.0
; SEQ ID NO 3636
; LENGTH: 218
; TYPE: PRT
; ORGANISM: Corynebacterium glutamicum
US-09-738-626-3636
```

```
Query Match      5.7%; Score 6; DB 9; Length 218;
Best Local Similarity 100.0%; Pred. No. 1.7e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

Qy 84 IVGYGR 89

Db 132 IVGYGR 137

## RESULT 23

```
US-09-938-803-4
; Sequence 4, Application US/09938803
; Patent No. US20020076762A1
; GENERAL INFORMATION:
; APPLICANT: Yue, Henry
; APPLICANT: Tang, Y. Tom
; APPLICANT: Lal, Preeti
; APPLICANT: Reddy, Roopa
; APPLICANT: Baughn, Mariah R.
; APPLICANT: Yang, Junning
; APPLICANT: Azimzai, Yalda
; TITLE OF INVENTION: FULL-LENGTH EXPRESSED GENETIC MARKERS
; FILE REFERENCE: PF-0695 US
; CURRENT APPLICATION NUMBER: US/09/938,803
; CURRENT FILING DATE: 2000-08-24
; PRIOR APPLICATION NUMBER: 09/311,894
; PRIOR FILING DATE: 1999-05-14
; NUMBER OF SEQ ID NOS: 40
; SOFTWARE: PERL Program
; SEQ ID NO 4
; LENGTH: 220
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; OTHER INFORMATION: incyte clone 1911421
US-09-938-803-4
```

```
Query Match      5.7%; Score 6; DB 10; Length 220;
Best Local Similarity 100.0%; Pred. No. 1.7e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

Qy 33 SARQR 38

Db 77 SARQR 82

## RESULT 24

```
US-09-881-752A-60
; Sequence 60, Application US/09881752A
; Patent No. US20020115078A1
; GENERAL INFORMATION:
; APPLICANT: Kleantous, Harold
; APPLICANT: Al-Garawi, Amal
; APPLICANT: Miller, Charles
; APPLICANT: Tomb, Jean-Francois
; APPLICANT: Oomen, Raymond P.
; TITLE OF INVENTION: Identification of Polynucleotides
; TITLE OF INVENTION: Encoding No. US20020115078A1 Helicobacter Polypeptides in the
; TITLE OF INVENTION: Genome
; FILE REFERENCE: 06132/041002
; CURRENT APPLICATION NUMBER: US/09/881,752A
; CURRENT FILING DATE: 2001-06-15
; PRIOR APPLICATION NUMBER: US 08/833,457
; PRIOR FILING DATE: 1997-04-01
; NUMBER OF SEQ ID NOS: 370
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 60
; LENGTH: 221
; TYPE: PRT
; ORGANISM: Helicobacter pylori
US-09-881-752A-60
```

```
Query Match      5.7%; Score 6; DB 10; Length 221;
Best Local Similarity 100.0%; Pred. No. 1.7e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

Qy 25 PGYMP 30

Db 172 PGYMP 177

## RESULT 25

```
US-09-864-761-35704
; Sequence 35704, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharon G.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
; FILE REFERENCE: Aecm1ca-X-1
; CURRENT APPLICATION NUMBER: US/09/864,761
; CURRENT FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/632,366
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: GB 24263,6
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US01/00666
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00664
```

```
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00669
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00665
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00668
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00663
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00662
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00661
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00670
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: US 09/608,408
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/774,203
; PRIOR FILING DATE: 2001-01-29
; NUMBER OF SEQ ID NOS: 49117
; SOFTWARE: Anomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 35704
; LENGTH: 255
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AF165423.1
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 0.96
; OTHER INFORMATION: EXPRESSED IN HBL100, SIGNAL = 0.95
; OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 0.98
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 1.2
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 1.1
; OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 1.5
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 1.5
; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 2.4
; OTHER INFORMATION: EST HUMAN HIT: BF345274.1, EVALU8 5.00e-42
; OTHER INFORMATION: SWISSPROT HIT: P35894, EVALU8 8.00e-62
US-09-864-761-35704

Query Match      5.7%; Score 6; DB 10; Length 255;
Best Local Similarity 100.0%; Pred. No. 2e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy      75 IPPPCT 80
      |||||
Db      16 IPPPCT 21

RESULT 26
US-09-815-242-14092
; Sequence 14092, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; TITLE OF INVENTION: Prokaryotes
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
```

```
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 14092
; LENGTH: 263
; TYPE: PRT
; ORGANISM: Salmonella typhi
US-09-815-242-14092

Query Match      5.7%; Score 6; DB 10; Length 263;
Best Local Similarity 100.0%; Pred. No. 2e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy      61 QSGIRV 66
      |||||
Db      175 QSGIRV 180

RESULT 27
US-09-808-665A-6
; Sequence 6, Application US/09808665A
; Publication No. US2003002822A1
; GENERAL INFORMATION:
; APPLICANT: David Michalovich
; APPLICANT: Trudy Rachel Doe
; APPLICANT: Philip David Hayes
; TITLE OF INVENTION: NOVEL COMPOUNDS
; FILE REFERENCE: GP-30036-C1
; CURRENT APPLICATION NUMBER: US/09/808,665A
; CURRENT FILING DATE: 2001-03-15
; PRIOR APPLICATION NUMBER: 09/160,762
; PRIOR FILING DATE: 1998-09-24
; PRIOR APPLICATION NUMBER: 98300693.3
; PRIOR FILING DATE: 1998-01-30
; PRIOR APPLICATION NUMBER: 9816423.9
; PRIOR FILING DATE: 1998-07-28
; PRIOR APPLICATION NUMBER: 9816676.2
; PRIOR FILING DATE: 1998-07-30
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 6
; LENGTH: 266
; TYPE: PRT
; ORGANISM: HOMO SAPIENS
; FEATURE:
; NAME/KEY: UNSURE
; LOCATION: (240)
; OTHER INFORMATION: OTHER INFORMATION: Partial amino acid sequence
US-09-808-665A-6

Query Match      5.7%; Score 6; DB 9; Length 266;
Best Local Similarity 100.0%; Pred. No. 2.1e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy      16 HOPSGV 21
      |||||
Db      254 HOPSGV 259

RESULT 28
US-09-975-139-1
; Sequence 1, Application US/09975139
; Patent No. US20020155460A1
; GENERAL INFORMATION:
; APPLICANT: Genencor International, Inc.
```

```
; APPLICANT: Schellenberger, Volker
; APPLICANT: Naki, Donald
; APPLICANT: Morrison, Thomas B.
; TITLE OF INVENTION: INFORMATION RICH LIBRARIES
; FILE REFERENCE: 23623-7060
; CURRENT APPLICATION NUMBER: US/09/975,139
; CURRENT FILING DATE: 2001-10-10
; PRIOR APPLICATION NUMBER: US 60/239,476
; PRIOR FILING DATE: 2000-10-10
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1
; LENGTH: 269
; TYPE: PRT
; ORGANISM: Bacillus lentus
; FEATURE:
; OTHER INFORMATION: Savinase - subtilisin protease
US-09-975-139-1

Query Match          5.7%; Score 6; DB 9; Length 269;
Best Local Similarity 100.0%; Pred. No. 2,1e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 19 SGVKVA 24
      |||||
      24 SGVKVA 29

Db 24 SGVKVA 29

RESULT 29
US-09-976-414-8
; Sequence 8, Application US/09976414
; Patent No. US20020155575A1
; GENERAL INFORMATION:
; APPLICANT: No. US20020155575A1regard-Madsen, Mads
; APPLICANT: Larsen, Line
; APPLICANT: Hansen, Peter
; TITLE OF INVENTION: Subtilase Variants
; FILE REFERENCE: 10081.200-US
; CURRENT APPLICATION NUMBER: US/09/976,414
; CURRENT FILING DATE: 2001-10-12
; NUMBER OF SEQ ID NOS: 8
; SOFTWARE: Patentin version 3.1
; SEQ ID NO 8
; LENGTH: 269
; TYPE: PRT
; ORGANISM: Bacillus
US-09-976-414-8

Query Match          5.7%; Score 6; DB 9; Length 269;
Best Local Similarity 100.0%; Pred. No. 2,1e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 19 SGVKVA 24
      |||||
      24 SGVKVA 29

Db 24 SGVKVA 29

RESULT 30
US-09-813-408-3
; Sequence 3, Application US/09813408
; Publication No. US20030049619A1
; GENERAL INFORMATION:
; APPLICANT: Delagrave, Simon
; APPLICANT: Matris, Barry
; TITLE OF INVENTION: Methods For The Synthesis Of Polynucleotides And Combinatorial Li
; FILE REFERENCE: HER0041
; CURRENT APPLICATION NUMBER: US/09/813,408
; CURRENT FILING DATE: 2001-03-21
; NUMBER OF SEQ ID NOS: 85
; SOFTWARE: Patentin version 3.0
; SEQ ID NO 3
; LENGTH: 269
```

```
; TYPE: PRT
; ORGANISM: Bacillus sp. KSM-K16
US-09-813-408-3

Query Match          5.7%; Score 6; DB 9; Length 269;
Best Local Similarity 100.0%; Pred. No. 2,1e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 19 SGVKVA 24
      |||||
      24 SGVKVA 29

Db 24 SGVKVA 29

RESULT 31
US-09-813-408-4
; Sequence 4, Application US/09813408
; Publication No. US20030049619A1
; GENERAL INFORMATION:
; APPLICANT: Delagrave, Simon
; APPLICANT: Matris, Barry
; TITLE OF INVENTION: Methods For The Synthesis Of Polynucleotides And Combinatorial I
; FILE REFERENCE: HER0041
; CURRENT APPLICATION NUMBER: US/09/813,408
; CURRENT FILING DATE: 2001-03-21
; NUMBER OF SEQ ID NOS: 85
; SOFTWARE: Patentin version 3.0
; SEQ ID NO 4
; LENGTH: 269
; TYPE: PRT
; ORGANISM: Bacillus lentus
; FEATURE:
; NAME/KEY: UNSURE
; LOCATION: (215)..(215)
; OTHER INFORMATION: X is unknown
US-09-813-408-4

Query Match          5.7%; Score 6; DB 9; Length 269;
Best Local Similarity 100.0%; Pred. No. 2,1e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 19 SGVKVA 24
      |||||
      24 SGVKVA 29

Db 24 SGVKVA 29

RESULT 32
US-09-813-408-5
; Sequence 5, Application US/09813408
; Publication No. US20030049619A1
; GENERAL INFORMATION:
; APPLICANT: Delagrave, Simon
; APPLICANT: Matris, Barry
; TITLE OF INVENTION: Methods For The Synthesis Of Polynucleotides And Combinatorial I
; FILE REFERENCE: HER0041
; CURRENT APPLICATION NUMBER: US/09/813,408
; CURRENT FILING DATE: 2001-03-21
; NUMBER OF SEQ ID NOS: 85
; SOFTWARE: Patentin version 3.0
; SEQ ID NO 5
; LENGTH: 269
; TYPE: PRT
; ORGANISM: Bacillus lentus
US-09-813-408-5

Query Match          5.7%; Score 6; DB 9; Length 269;
Best Local Similarity 100.0%; Pred. No. 2,1e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 19 SGVKVA 24
      |||||
      24 SGVKVA 29

Db 24 SGVKVA 29
```

```
RESULT 33
US-09-837-235-16
; Sequence 16, Application US/09837235
; Patent No. US20020061549A1
; GENERAL INFORMATION:
; APPLICANT: Marshall, Christopher
; APPLICANT: Hoffman, Alexander
; APPLICANT: Errico, Joseph
; APPLICANT: Marshall, Paul
; TITLE OF INVENTION: STABILIZED PROTEINS
; FILE REFERENCE: 9725-005-999
; CURRENT APPLICATION NUMBER: US/09/837,235
; CURRENT FILING DATE: 2001-04-18
; PRIOR APPLICATION NUMBER: PCT/US00/28595
; PRIOR FILING DATE: 2000-10-16
; PRIOR APPLICATION NUMBER: 60/159,763
; PRIOR FILING DATE: 1999-10-15
; NUMBER OF SEQ ID NOS: 42
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 16
; LENGTH: 269
; TYPE: PRT
; ORGANISM: Bacillus subtilis
US-09-837-235-16

Query Match      5.7%; Score 6; DB 10; Length 269;
Best Local Similarity 100.0%; Pred. No. 2.1e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      19 SGVKVA 24
      |||||
Db      24 SGVKVA 29

RESULT 34
US-09-060-854B-6
; Sequence 6, Application US/09060854B
; Patent No. US20020081703A1
; GENERAL INFORMATION:
; APPLICANT: Estell, David Aaron
; TITLE OF INVENTION: Human Protease and Use of Such Protease for Pharmaceutical
; TITLE OF INVENTION: Applications and for Reducing the Allergenicity of No. US2002008
; TITLE OF INVENTION: Proteins
; FILE REFERENCE: GC532
; CURRENT APPLICATION NUMBER: US/09/060,854B
; CURRENT FILING DATE: 1998-04-15
; NUMBER OF SEQ ID NOS: 7
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 6
; LENGTH: 269
; TYPE: PRT
; ORGANISM: B. lentus
US-09-060-854B-6

Query Match      5.7%; Score 6; DB 10; Length 269;
Best Local Similarity 100.0%; Pred. No. 2.1e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      19 SGVKVA 24
      |||||
Db      24 SGVKVA 29

RESULT 35
US-10-075-907-1
; Sequence 1, Application US/10075907
; Patent No. US20020127695A1
; GENERAL INFORMATION:
; APPLICANT: Genencor International, Inc.
; APPLICANT: Davis, Benjamin G.
; APPLICANT: Jones, Bryan J.
```

```
; APPLICANT: Bott, Richard R.
; TITLE OF INVENTION: CHEMICALLY MODIFIED ENZYMES WITH
; TITLE OF INVENTION: MULTIPLE CHARGED VARIANTS
; FILE REFERENCE: 23623-7039
; CURRENT APPLICATION NUMBER: US/10/075,907
; CURRENT FILING DATE: 2002-02-12
; PRIOR APPLICATION NUMBER: US 60/113,130
; PRIOR FILING DATE: 1998-12-21
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1
; LENGTH: 269
; TYPE: PRT
; ORGANISM: Bacillus lentus
; FEATURE:
; OTHER INFORMATION: Subtilisin SBL
US-10-075-907-1

Query Match      5.7%; Score 6; DB 12; Length 269;
Best Local Similarity 100.0%; Pred. No. 2.1e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      19 SGVKVA 24
      |||||
Db      24 SGVKVA 29

RESULT 36
US-10-075-895-1
; Sequence 1, Application US/10075895
; Patent No. US20020137177A1
; GENERAL INFORMATION:
; APPLICANT: Genencor International, Inc.
; APPLICANT: Jones, Bryan J.
; TITLE OF INVENTION: MODIFIED ENZYMES AND THEIR USE FOR
; TITLE OF INVENTION: PEPTIDE SYNTHESIS
; FILE REFERENCE: 23623-7013
; CURRENT APPLICATION NUMBER: US/10/075,895
; CURRENT FILING DATE: 2002-02-13
; PRIOR APPLICATION NUMBER: US 60/072,265
; PRIOR FILING DATE: 1998-01-23
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1
; LENGTH: 269
; TYPE: PRT
; ORGANISM: Bacillus lentus
; FEATURE:
; OTHER INFORMATION: Subtilisin SBL
US-10-075-895-1

Query Match      5.7%; Score 6; DB 12; Length 269;
Best Local Similarity 100.0%; Pred. No. 2.1e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      19 SGVKVA 24
      |||||
Db      24 SGVKVA 29

RESULT 37
US-09-813-408-2
; Sequence 2, Application US/09813408
; Publication No. US20030049619A1
; GENERAL INFORMATION:
; APPLICANT: Delagrave, Simon
; APPLICANT: Maris, Barry
; TITLE OF INVENTION: Methods For The Synthesis Of Polynucleotides And Combinatorial L
; TITLE OF INVENTION: OF Polynucleotides
; FILE REFERENCE: HR0041
; CURRENT APPLICATION NUMBER: US/09/813,408
; CURRENT FILING DATE: 2001-03-21
; NUMBER OF SEQ ID NOS: 85
```



```
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 2
; LENGTH: 271
; TYPE: PRT
; ORGANISM: Bacillus lentus
US-09-813-408-2
```

```
Query Match          5.7%; Score 6; DB 9; Length 271;
Best Local Similarity 100.0%; Pred. No. 2.1e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY      19 SGVXVA 24
      |||||
Db       24 SGVXVA 29
```

```
RESULT 38
US-09-815-242-10241
; Sequence 10241, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari L.
; APPLICANT: Zyckind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 10241
; LENGTH: 285
; TYPE: PRT
; ORGANISM: Escherichia coli
US-09-815-242-10241

Query Match          5.7%; Score 6; DB 10; Length 285;
Best Local Similarity 100.0%; Pred. No. 2.2e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY      61 QSGIRV 66
      |||||
Db       197 QSGIRV 202
```

```
RESULT 39
US-09-738-626-6072
; Sequence 6072, Application US/09738626
; Publication No. US20020197605A1
; GENERAL INFORMATION:
; APPLICANT: NAKAGAWA, SATOSHI
; APPLICANT: MIZOGUCHI, HIROSHI
; APPLICANT: ANDO, SEIKO
```

```
; APPLICANT: HAYASHI, MIKIRO
; APPLICANT: OCHIAI, KEIKO
; APPLICANT: YOKOI, HARUHIKO
; APPLICANT: TATEISHI, NAOKO
; APPLICANT: SENOH, AKIHIRO
; APPLICANT: IKEDA, MASATO
; APPLICANT: OZAKI, AKIO
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-125
; CURRENT APPLICATION NUMBER: US/09/738,626
; CURRENT FILING DATE: 2000-12-18
; PRIOR APPLICATION NUMBER: JP 99/377484
; PRIOR FILING DATE: 1999-12-16
; PRIOR APPLICATION NUMBER: JP 00/159162
; PRIOR FILING DATE: 2000-04-07
; PRIOR APPLICATION NUMBER: JP 00/280988
; PRIOR FILING DATE: 2000-08-03
; NUMBER OF SEQ ID NOS: 7059
; SOFTWARE: PatentIn ver. 3.0
; SEQ ID NO 6072
; LENGTH: 325
; TYPE: PRT
; ORGANISM: Corynebacterium glutamicum
US-09-738-626-6072
```

```
Query Match          5.7%; Score 6; DB 9; Length 325;
Best Local Similarity 100.0%; Pred. No. 2.5e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY      45 SGFDIA 50
      |||||
Db       202 SGFDIA 207
```

```
RESULT 40
US-09-252-088-35
; Sequence 35, Application US/09252088
; Publication No. US20030031682A1
; GENERAL INFORMATION:
; APPLICANT: BRODEUR, Bernard R.
; APPLICANT: RIOUX, Clement
; APPLICANT: BOYER, Martine
; APPLICANT: CHARLEBOIS, Isabelle
; APPLICANT: HAMEL, Jose
; APPLICANT: MARTIN, Denis
; TITLE OF INVENTION: NOVEL GROUP B STREPTOCOCCUS ANTIGENS
; FILE REFERENCE: 8331-9002
; CURRENT APPLICATION NUMBER: US/09/252,088
; CURRENT FILING DATE: 1999-02-18
; EARLIER APPLICATION NUMBER: US/60/075,425
; EARLIER FILING DATE: 1998-02-20
; NUMBER OF SEQ ID NOS: 44
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 35
; LENGTH: 326
; TYPE: PRT
; ORGANISM: group B streptococcus
US-09-252-088-35
```

```
Query Match          5.7%; Score 6; DB 9; Length 326;
Best Local Similarity 100.0%; Pred. No. 2.5e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY      11 QOGIHH 16
      |||||
Db       162 QOGIHH 167
```

```
Search completed: April 9, 2003, 13:27:14
Job time : 60.0984 secs
```

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.4.p5.4578  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 9, 2003, 12:59:07 ; Search time 88.623 Seconds  
(without alignments)  
159.378 Million cell updates/sec

Title: US-10-020-441-2

Perfect score: 106

Sequence: 1 VSFPLTNRNGDQGIHQPSG.....YGRDDNRDPRKNGILKK 106

Scoring table:

Gapop 60.0 , Gapext 60.0

Searched: 908470 reqs, 133250620 .residues

Word size : 0

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 1000 summaries

Database :

\_A\_Geneseq\_101002:\*

1: /SIDs2/gcgdata/geneseq/geneseq-emb1/AA1980.DAT:\*  
2: /SIDs2/gcgdata/geneseq/geneseq-emb1/AA1981.DAT:\*  
3: /SIDs2/gcgdata/geneseq/geneseq-emb1/AA1982.DAT:\*  
4: /SIDs2/gcgdata/geneseq/geneseq-emb1/AA1983.DAT:\*  
5: /SIDs2/gcgdata/geneseq/geneseq-emb1/AA1984.DAT:\*  
6: /SIDs2/gcgdata/geneseq/geneseq-emb1/AA1985.DAT:\*  
7: /SIDs2/gcgdata/geneseq/geneseq-emb1/AA1986.DAT:\*  
8: /SIDs2/gcgdata/geneseq/geneseq-emb1/AA1987.DAT:\*  
9: /SIDs2/gcgdata/geneseq/geneseq-emb1/AA1988.DAT:\*  
10: /SIDs2/gcgdata/geneseq/geneseq-emb1/AA1989.DAT:\*  
11: /SIDs2/gcgdata/geneseq/geneseq-emb1/AA1990.DAT:\*  
12: /SIDs2/gcgdata/geneseq/geneseq-emb1/AA1991.DAT:\*  
13: /SIDs2/gcgdata/geneseq/geneseq-emb1/AA1992.DAT:\*  
14: /SIDs2/gcgdata/geneseq/geneseq-emb1/AA1993.DAT:\*  
15: /SIDs2/gcgdata/geneseq/geneseq-emb1/AA1994.DAT:\*  
16: /SIDs2/gcgdata/geneseq/geneseq-emb1/AA1995.DAT:\*  
17: /SIDs2/gcgdata/geneseq/geneseq-emb1/AA1996.DAT:\*  
18: /SIDs2/gcgdata/geneseq/geneseq-emb1/AA1997.DAT:\*  
19: /SIDs2/gcgdata/geneseq/geneseq-emb1/AA1998.DAT:\*  
20: /SIDs2/gcgdata/geneseq/geneseq-emb1/AA1999.DAT:\*  
21: /SIDs2/gcgdata/geneseq/geneseq-emb1/AA2000.DAT:\*  
22: /SIDs2/gcgdata/geneseq/geneseq-emb1/AA2001.DAT:\*  
23: /SIDs2/gcgdata/geneseq/geneseq-emb1/AA2002.DAT:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	106	100.0	106	21	AAV67587
2	16	15.1	16	21	AAV67588
3	10	9.4	10	16	AAW21482
4	9	8.5	11	16	AAW21484
5	7	6.6	150	19	AAV85951
6	7	6.6	186	21	AAAG14293
7	7	6.6	341	17	AAAR8326
8	7	6.6	356	15	AAAR63225
9	7	6.6	356	22	AAAB6207
10	7	6.6	359	23	AAU11996

11	7	6.6	402	22	ABG04674	Novel human diagno
12	7	6.6	725	21	AAAB14350	Human partial myel
13	7	6.6	810	21	AAAG48794	Arabidopsis thalia
14	7	6.6	829	21	AAAB23339	Human ORFX ORF2103
15	7	6.6	838	22	AAAB64668	Drosophila melanog
16	7	6.6	2724	22	AAU08681	Human FCTR3J polyp
17	7	6.6	2733	22	AAU08680	Human FCTR3J polyp
18	6	5.7	6	16	AAW21483	Schistosoma elasta
19	6	5.7	10	22	AAAG5984	Human complementar
20	6	5.7	15	21	AAV66990	B. lentus protease
21	6	5.7	15	21	AAV66991	B. lentus protease
22	6	5.7	15	21	AAV66992	B. lentus protease
23	6	5.7	15	21	AAV54627	B. lentus protease
24	6	5.7	15	21	AAV54628	B. lentus protease
25	6	5.7	15	21	AAV54629	B. lentus protease
26	6	5.7	15	22	AAU38529	Bacillus lentus sa
27	6	5.7	15	22	AAU38530	Bacillus lentus sa
28	6	5.7	15	22	AAU38531	Bacillus lentus sa
29	6	5.7	21	20	AAV29824	pneumolysin (263-8
30	6	5.7	30	22	ABR37761	Peptide #5267 enco
31	6	5.7	30	23	ABG40652	Human peptide enco
32	6	5.7	34	22	ABR22985	Protein #4984 enco
33	6	5.7	34	22	AAV50227	Human immune/haema
34	6	5.7	39	21	AAE04671	Haem protein relat
35	6	5.7	51	21	AAAG03372	Human secreted pro
36	6	5.7	54	22	AAU41874	Apomoglobin BETA
37	6	5.7	60	17	AAAB95436	Ppu DnaB intein am
38	6	5.7	61	21	AAAB23748	Protonibacterium
39	6	5.7	62	22	AAU64135	Novel human secret
40	6	5.7	62	22	AAU32172	Human ORF581 prote
41	6	5.7	64	23	AAU42617	Protonibacterium
42	6	5.7	66	22	AAU42617	Human polypeptide
43	6	5.7	66	23	AAAB6955	Protonibacterium
44	6	5.7	67	22	AAAB6956	Human polypeptide
45	6	5.7	67	22	AAAB6957	Human immune/haema
46	6	5.7	70	22	AAAB17169	Staphylococcus epi
47	6	5.7	72	22	AAAB1427	Human hematology
48	6	5.7	75	22	AAAB1691	Human hematology
49	6	5.7	77	22	AAAB69474	Drosophila melanog
50	6	5.7	78	22	AAAB69474	Human immune/haema
51	6	5.7	74	22	AAAB3244	Protonibacterium
52	6	5.7	74	23	AAAB39448	Novel human diagno
53	6	5.7	75	22	AAAB1427	Human secreted pro
54	6	5.7	76	22	AAAB1691	Human polypeptide
55	6	5.7	76	22	AAAB69474	Protonibacterium
56	6	5.7	84	22	AAAB5277	Human polypeptide
57	6	5.7	84	22	AAAB5277	Human immune/haema
58	6	5.7	85	22	AAAB5277	Protonibacterium
59	6	5.7	88	22	AAAB5277	Novel human diagno
60	6	5.7	88	22	AAAB5277	Human secreted pro
61	6	5.7	89	22	AAAB5277	Human polypeptide
62	6	5.7	91	23	AAAB5277	Protonibacterium
63	6	5.7	93	22	AAAB5277	Human ORFX protei
64	6	5.7	95	19	AAAB5277	Human digestive sy
65	6	5.7	95	19	AAAB5277	Conus quercinus co
66	6	5.7	95	19	AAAB5277	Conus quercinus co
67	6	5.7	95	19	AAAB5277	Arabidopsis thalia
68	6	5.7	95	19	AAAB5277	Protonibacterium
69	6	5.7	98	22	AAAB5277	Amino acid sequenc
70	6	5.7	98	22	AAAB5277	Propeptide of cono
71	6	5.7	99	22	AAAB5277	Propeptide of cono
72	6	5.7	101	23	AAAB5277	Drosophila melanog
73	6	5.7	101	23	AAAB5277	Human ORF2167 prot
74	6	5.7	104	23	AAAB5277	Human alkylation D
75	6	5.7	105	22	AAAB5277	Protonibacterium
76	6	5.7	106	22	AAAB5277	Human polypeptide
77	6	5.7	107	21	AAAB5277	Human secreted pro
78	6	5.7	110	22	AAAB5277	Protonibacterium
79	6	5.7	111	22	AAAB5277	Human polypeptide
80	6	5.7	111	22	AAAB5277	Human secreted pro
81	6	5.7	113	21	AAAB5277	Protonibacterium
82	6	5.7	114	23	AAAB5277	Human ORF957 prote
83	6	5.7	119	21	AAAB5277	Lung cancer associ
84	6	5.7	119	22	AAAB5277	Novel human diagno

84	6	5.7	124	22	AA001192	Human polypeptide
85	6	5.7	127	22	AAU50232	Propionibacterium
86	6	5.7	127	23	ABP01823	Human ORFX protein
87	6	5.7	128	22	ABR30411	Peptide #3062 enco
88	6	5.7	128	22	ABR21010	Protein #3009 enco
89	6	5.7	128	22	AAW56398	Human brain expres
90	6	5.7	128	22	AAW68978	Human bone marrow
91	6	5.7	128	22	AAW04313	Peptide #2995 enco
92	6	5.7	129	21	AAW45005	zebra mays protein f
93	6	5.7	130	21	AAW44026	Arabidopsis thalia
94	6	5.7	144	22	AAW91826	C glutamicum prote
95	6	5.7	144	22	AAW80201	Corynebacterium gl
96	6	5.7	145	9	AAW81003	Sequence encoding
97	6	5.7	146	22	ABR28111	Human peptide #762
98	6	5.7	146	22	ABR33286	Peptide #792 enco
99	6	5.7	146	22	ABR18747	Protein #746 enco
100	6	5.7	146	22	AAW54076	Human brain expres
101	6	5.7	146	22	AAW66466	Human bone marrow
102	6	5.7	146	22	AAW14338	Peptide #772 enco
103	6	5.7	146	22	AAW26750	Peptide #787 enco
104	6	5.7	146	22	AAW02067	Peptide #749 enco
105	6	5.7	146	22	ABG36117	Human peptide enco
106	6	5.7	146	23	ABP05051	Human ORFX protein
107	6	5.7	149	21	AAW55975	Arabidopsis thalia
108	6	5.7	151	21	AAW55974	Arabidopsis thalia
109	6	5.7	152	21	AAW44025	Arabidopsis thalia
110	6	5.7	154	19	AAW62270	Modified myoglobin
111	6	5.7	154	19	AAW62271	Modified myoglobin
112	6	5.7	154	19	AAW62272	Modified myoglobin
113	6	5.7	154	19	AAW62273	Modified myoglobin
114	6	5.7	154	19	AAW29739	Modified myoglobin
115	6	5.7	154	19	AAW29740	Modified myoglobin
116	6	5.7	154	19	AAW29741	Modified myoglobin
117	6	5.7	154	19	AAW29742	Modified myoglobin
118	6	5.7	154	19	AAW29743	Modified myoglobin
119	6	5.7	154	19	AAW29744	Modified myoglobin
120	6	5.7	154	21	AAW44024	Arabidopsis thalia
121	6	5.7	163	23	ABG03725	Novel human diagno
122	6	5.7	167	23	ABR89382	Human polypeptide
123	6	5.7	168	21	AAW29469	Cucumber xylen sap
124	6	5.7	170	20	AAW29298	Protein encoded by
125	6	5.7	172	22	AAU10059	fatly acid-binding
126	6	5.7	177	22	AAU33668	Pseudomonas aerugi
127	6	5.7	180	23	ABW54862	Lactococcus lactis
128	6	5.7	191	22	ABG28718	Novel human diagno
129	6	5.7	194	22	AAU31497	Novel human secret
130	6	5.7	196	22	AAU29458	Human G protein-co
131	6	5.7	196	23	ABG60746	Novel G protein co
132	6	5.7	198	23	ABP26122	Streptococcus poly
133	6	5.7	202	22	ABG04708	Novel human diagno
134	6	5.7	204	20	AAW76596	Human ovarian tumo
135	6	5.7	206	22	AAW83429	Human ions protein
136	6	5.7	215	16	AAW64202	Monoclonal antibod
137	6	5.7	218	22	AAW89882	C glutamicum prote
138	6	5.7	220	23	AAW36582	Human FLEXHR-4 pro
139	6	5.7	220	23	AAO21540	Human ribonuclease
140	6	5.7	221	19	AAW98230	H. pylori GHPO 157
141	6	5.7	223	19	AAW11061	H. pylori ORF 09ze
142	6	5.7	223	19	ABW54143	Lactococcus lactis
143	6	5.7	228	22	ABG10401	Novel human diagno
144	6	5.7	230	21	AAW28192	Arabidopsis thalia
145	6	5.7	230	21	AAW54986	Arabidopsis thalia
146	6	5.7	231	22	ABG04706	Novel human diagno
147	6	5.7	233	22	ABW71644	Drosophila melano
148	6	5.7	233	23	AAO18241	Streptomyces coeli
149	6	5.7	239	22	ABG24417	Novel human diagno
150	6	5.7	239	22	ABW58719	Novel human diagno
151	6	5.7	248	22	ABW58495	Drosophila melano
152	6	5.7	253	23	AAW52197	Mouse degranulatio
153	6	5.7	255	22	ABW29818	Peptide #2469 enco
154	6	5.7	255	22	ABW20406	Protein #2405 enco
155	6	5.7	255	22	AAW55806	Human brain expres
156	6	5.7	255	22	AAW68180	Human bone marrow
157	6	5.7	255	22	AAW16002	Peptide #2436 enco
158	6	5.7	255	22	AAW28504	Peptide #2541 enco
159	6	5.7	255	22	AAW03737	Peptide #2419 enco
160	6	5.7	257	21	AAW28191	Arabidopsis thalia
161	6	5.7	263	22	AAW38499	Salmonella typhi c
162	6	5.7	263	20	AAW17700	B. lantus subtilis
163	6	5.7	265	20	AAW16768	Bacillus lentus su
164	6	5.7	265	20	AAW16768	Bacillus lentus su
165	6	5.7	266	20	AAW23875	Protein encoded by
166	6	5.7	268	19	AAW62227	Subtilase BSAVI f
167	6	5.7	269	10	AAW90374	Subtilisin 309 fro
168	6	5.7	269	12	AAW11390	Pre-pro alkaline p
169	6	5.7	269	12	AAW10444	S004 mutant of the
170	6	5.7	269	12	AAW10445	S014 mutant of the
171	6	5.7	269	12	AAW10446	S020 mutant of the
172	6	5.7	269	12	AAW10447	S204 mutant of the
173	6	5.7	269	12	AAW10448	S224 mutant of the
174	6	5.7	269	12	AAW10449	S234 mutant of the
175	6	5.7	269	12	AAW10449	N18K mature highly
176	6	5.7	269	12	AAW11388	N42R mature highly
177	6	5.7	269	12	AAW11389	O57R mature highly
178	6	5.7	269	12	AAW11390	A96R mature highly
179	6	5.7	269	12	AAW11391	O107R mature highl
180	6	5.7	269	12	AAW11392	N14R mature highl
181	6	5.7	269	12	AAW11393	N15R mature highl
182	6	5.7	269	12	AAW11394	O135R mature highl
183	6	5.7	269	12	AAW11395	N138R mature highl
184	6	5.7	269	12	AAW11396	A166R mature highl
185	6	5.7	269	12	AAW11397	V238R mature highl
186	6	5.7	269	12	AAW11398	N258R mature highl
187	6	5.7	269	12	AAW11399	S259K mature highl
188	6	5.7	269	12	AAW11400	A266R mature highl
189	6	5.7	269	13	AAW28336	Subtilisin 309 (wi
190	6	5.7	269	13	AAW32321	Subtilisin 309 mut
191	6	5.7	269	13	AAW28378	Subtilisin 309 S
192	6	5.7	269	13	AAW29594	Mutant subtilisin
193	6	5.7	269	13	AAW29595	Mutant subtilisin
194	6	5.7	269	13	AAW28381	Subtilisin proteas
195	6	5.7	269	13	AAW25541	Mutant subtilisin
196	6	5.7	269	13	AAW28382	Mutant subtilisin
197	6	5.7	269	13	AAW25538	Subtilisin proteas
198	6	5.7	269	13	AAW25542	Subtilisin proteas
199	6	5.7	269	13	AAW25543	Subtilisin proteas
200	6	5.7	269	13	AAW25544	Subtilisin proteas
201	6	5.7	269	13	AAW25545	Subtilisin proteas
202	6	5.7	269	13	AAW25546	Subtilisin proteas
203	6	5.7	269	13	AAW25547	Subtilisin proteas
204	6	5.7	269	13	AAW25548	Subtilisin proteas
205	6	5.7	269	13	AAW25549	Subtilisin proteas
206	6	5.7	269	13	AAW25550	Subtilisin proteas
207	6	5.7	269	13	AAW25551	Subtilisin proteas
208	6	5.7	269	13	AAW25552	Subtilisin proteas
209	6	5.7	269	13	AAW28383	Mutant subtilisin
210	6	5.7	269	13	AAW28384	Mutant subtilisin
211	6	5.7	269	13	AAW28385	Mutant subtilisin
212	6	5.7	269	13	AAW28386	Mutant subtilisin
213	6	5.7	269	13	AAW28387	Mutant subtilisin
214	6	5.7	269	13	AAW28388	Mutant subtilisin
215	6	5.7	269	13	AAW28389	Mutant subtilisin
216	6	5.7	269	13	AAW28390	Mutant subtilisin
217	6	5.7	269	13	AAW28393	Ref. subtilisin-309 - c
218	6	5.7	269	13	AAW31430	Ref. subtilisin fo
219	6	5.7	269	13	AAW30114	BLAP (H243A). Bac
220	6	5.7	269	13	AAW30091	BLAP (S37, V4I, A1
221	6	5.7	269	13	AAW30092	BLAP (S37, A188P,
222	6	5.7	269	13	AAW30093	BLAP (V4I, A188P,
223	6	5.7	269	13	AAW30094	BLAP (S139Y, A188P
224	6	5.7	269	13	AAW30095	BLAP (S130T, S139Y
225	6	5.7	269	13	AAW30096	BLAP (A188P, V193M
226	6	5.7	269	13	AAW30097	BLAP (S37, A188P,
227	6	5.7	269	13	AAW30098	BLAP (S157I). Bac
228	6	5.7	269	13	AAW30099	BLAP (A188P, V139M
229	6	5.7	269	13	AAW30100	BLAP (A188P). Bac

230	6	5.7	269	13	AAK30101	BLAP (S3T, V4I, A1	303	6	5.7	269	15	AAK46320	Subtilisin 309 Ser
231	6	5.7	269	13	AAK30102	BLAP (V193M). Bac	304	6	5.7	269	15	AAK46321	PB92 serine protea
232	6	5.7	269	13	AAK30103	BLAP (S104T). Bac	305	6	5.7	269	15	AAK46322	Subtilisin 309 Ser
233	6	5.7	269	13	AAK30104	BLAP (T69V). Bac	306	6	5.7	269	15	AAK46323	PB92 serine protea
234	6	5.7	269	13	AAK30105	BLAP (V4I, A188P,	307	6	5.7	269	15	AAK46324	Subtilisin 309 Ser
235	6	5.7	269	13	AAK30106	BLAP (A224V). Bac	308	6	5.7	269	15	AAK46325	PB92 serine protea
236	6	5.7	269	13	AAK30107	BLAP (V199I). Bac	309	6	5.7	269	15	AAK46326	Subtilisin 309 Ser
237	6	5.7	269	13	AAK30108	BLAP (V4I). Bac1	310	6	5.7	269	15	AAK46327	PB92 serine protea
238	6	5.7	269	13	AAK30109	BLAP (S3T). Bac1	311	6	5.7	269	15	AAK46328	Subtilisin 309 Ser
239	6	5.7	269	13	AAK30110	BLAP (S139Y). Bac	312	6	5.7	269	15	AAK46329	PB92 serine protea
240	6	5.7	269	13	AAK30111	BLAP (N242A). Bac	313	6	5.7	269	15	AAK46330	Subtilisin 309 Ser
241	6	5.7	269	13	AAK30112	BLAP (S236T). Bac	314	6	5.7	269	15	AAK46331	PB92 serine protea
242	6	5.7	269	13	AAK30113	BLAP (S36A). Bac1	315	6	5.7	269	15	AAK46332	Subtilisin 309 Ser
243	6	5.7	269	13	AAK30115	BLAP (A101T). Bac	316	6	5.7	269	15	AAK46333	PB92 serine protea
244	6	5.7	269	13	AAK30116	BLAP (S236A). Bac	317	6	5.7	269	15	AAK46334	Subtilisin 309 Ser
245	6	5.7	269	13	AAK30117	BLAP (B87R). Bac1	318	6	5.7	269	15	AAK46335	PB92 serine protea
246	6	5.7	269	13	AAK30118	BLAP (N114S). Bac	319	6	5.7	269	15	AAK46336	Subtilisin 309 Ser
247	6	5.7	269	13	AAK30119	BLAP (A47W). Bac1	320	6	5.7	269	15	AAK46337	PB92 serine protea
248	6	5.7	269	13	AAK30120	BLAP (A120S). Bac	321	6	5.7	269	15	AAK46338	Subtilisin 309 Ser
249	6	5.7	269	13	AAK30121	BLAP (T56V). Bac1	322	6	5.7	269	15	AAK46339	PB92 serine protea
250	6	5.7	269	13	AAK30122	BLAP (A120V). Bac	323	6	5.7	269	15	AAK46340	Subtilisin 309 Ser
251	6	5.7	269	13	AAK30123	BLAP (G205V). Bac	324	6	5.7	269	15	AAK46341	PB92 serine protea
252	6	5.7	269	13	AAK30124	BLAP (S130A). Bac	325	6	5.7	269	15	AAK46342	Subtilisin 309 Ser
253	6	5.7	269	13	AAK30125	BLAP (S130T). Bac	326	6	5.7	269	15	AAK46343	PB92 serine protea
254	6	5.7	269	13	AAK30126	BLAP (A96I). Bac1	327	6	5.7	269	15	AAK46344	Subtilisin 309 Ser
255	6	5.7	269	13	AAK30127	BLAP (S104T, S139Y	328	6	5.7	269	15	AAK46345	PB92 serine protea
256	6	5.7	269	13	AAK30128	BLAP (S139A). Bac	329	6	5.7	269	15	AAK46346	Subtilisin 309 Ser
257	6	5.7	269	13	AAK30129	BLAP (S142T). Bac	330	6	5.7	269	15	AAK46347	PB92 serine protea
258	6	5.7	269	13	AAK30130	BLAP (S139T). Bac	331	6	5.7	269	15	AAK46348	Subtilisin 309 Ser
259	6	5.7	269	13	AAK30131	BLAP (I102W). Bac	332	6	5.7	269	15	AAK46349	PB92 serine protea
260	6	5.7	269	13	AAK30132	BLAP (A96N). Bac1	333	6	5.7	269	15	AAK46350	Subtilisin 309 Ser
261	6	5.7	269	13	AAK30133	BLAP (N42F). Bac1	334	6	5.7	269	15	AAK46351	PB92 serine protea
262	6	5.7	269	13	AAK30134	BLAP (S142A). Bac	335	6	5.7	269	15	AAK46352	Subtilisin 309 Ser
263	6	5.7	269	13	AAK30135	BLAP (H118P). Bac	336	6	5.7	269	15	AAK46353	PB92 serine protea
264	6	5.7	269	13	AAK30136	BLAP (N237A). Bac	337	6	5.7	269	15	AAK46354	Subtilisin 309 Ser
265	6	5.7	269	13	AAK30137	BLAP (N255P). Bac	338	6	5.7	269	15	AAK46355	PB92 serine protea
266	6	5.7	269	13	AAK30138	BLAP (T141W, N237A	339	6	5.7	269	15	AAK46356	Subtilisin 309 Ser
267	6	5.7	269	13	AAK30139	BLAP (T268V). Bac	340	6	5.7	269	15	AAK46357	PB92 serine protea
268	6	5.7	269	13	AAK30140	BLAP (K229W). Bac	341	6	5.7	269	15	AAK46358	Subtilisin 309 Ser
269	6	5.7	269	13	AAK30141	BLAP (T141W). Bac	342	6	5.7	269	15	AAK46359	PB92 serine protea
270	6	5.7	269	13	AAK30142	BLAP (wildtype). Bac	343	6	5.7	269	15	AAK46360	Subtilisin 309 Ser
271	6	5.7	269	14	AAK3047	Mutant PB92 serine	344	6	5.7	269	15	AAK46361	PB92 serine protea
272	6	5.7	269	14	AAK3048	Mutant PB92 serine	345	6	5.7	269	15	AAK46362	Subtilisin 309 Ser
273	6	5.7	269	14	AAK3049	Mutant PB92 serine	346	6	5.7	269	15	AAK46363	PB92 serine protea
274	6	5.7	269	14	AAK3050	Mutant PB92 serine	347	6	5.7	269	15	AAK46364	Subtilisin 309 Ser
275	6	5.7	269	14	AAK3051	Mutant PB92 serine	348	6	5.7	269	15	AAK46365	PB92 serine protea
276	6	5.7	269	14	AAK3052	Mutant PB92 serine	349	6	5.7	269	15	AAK46366	Subtilisin 309 Ser
277	6	5.7	269	14	AAK3053	Mutant PB92 serine	350	6	5.7	269	15	AAK46367	PB92 serine protea
278	6	5.7	269	14	AAK3054	Mutant PB92 serine	351	6	5.7	269	15	AAK46368	Subtilisin 309 Ser
279	6	5.7	269	14	AAK3055	Mutant PB92 serine	352	6	5.7	269	15	AAK46369	PB92 serine protea
280	6	5.7	269	14	AAK3056	Mutant PB92 serine	353	6	5.7	269	15	AAK46370	Subtilisin 309 Ser
281	6	5.7	269	14	AAK3057	Mutant PB92 serine	354	6	5.7	269	15	AAK46371	PB92 serine protea
282	6	5.7	269	14	AAK3058	Mutant PB92 serine	355	6	5.7	269	15	AAK46372	Subtilisin 309 Ser
283	6	5.7	269	14	AAK3059	Mutant PB92 serine	356	6	5.7	269	15	AAK46373	PB92 serine protea
284	6	5.7	269	15	AAK46301	WT PB92 serine pro	357	6	5.7	269	15	AAK46374	Subtilisin 309 Ser
285	6	5.7	269	15	AAK46302	WT PB92 serine pro	358	6	5.7	269	15	AAK46375	PB92 serine protea
286	6	5.7	269	15	AAK46303	PB92 serine protea	359	6	5.7	269	15	AAK46376	Subtilisin 309 Ser
287	6	5.7	269	15	AAK46304	Subtilisin 309 Ser	360	6	5.7	269	15	AAK46377	PB92 serine protea
288	6	5.7	269	15	AAK46305	PB92 serine protea	361	6	5.7	269	15	AAK46378	Subtilisin 309 Ser
289	6	5.7	269	15	AAK46306	Subtilisin 309 Ser	362	6	5.7	269	15	AAK46379	PB92 serine protea
290	6	5.7	269	15	AAK46307	PB92 serine protea	363	6	5.7	269	15	AAK46380	Subtilisin 309 Ser
291	6	5.7	269	15	AAK46308	Subtilisin 309 Ser	364	6	5.7	269	15	AAK46381	PB92 serine protea
292	6	5.7	269	15	AAK46309	PB92 serine protea	365	6	5.7	269	15	AAK46382	Subtilisin 309 Ser
293	6	5.7	269	15	AAK46310	Subtilisin 309 Ser	366	6	5.7	269	15	AAK46383	PB92 serine protea
294	6	5.7	269	15	AAK46311	PB92 serine protea	367	6	5.7	269	15	AAK46384	Subtilisin 309 Ser
295	6	5.7	269	15	AAK46312	PB92 serine protea	368	6	5.7	269	15	AAK46385	PB92 serine protea
296	6	5.7	269	15	AAK46313	PB92 serine protea	369	6	5.7	269	15	AAK46386	Subtilisin 309 Ser
297	6	5.7	269	15	AAK46314	Subtilisin 309 Ser	370	6	5.7	269	15	AAK46387	PB92 serine protea
298	6	5.7	269	15	AAK46315	PB92 serine protea	371	6	5.7	269	15	AAK46388	Subtilisin 309 Ser
299	6	5.7	269	15	AAK46316	Subtilisin 309 Ser	372	6	5.7	269	15	AAK46389	PB92 serine protea
300	6	5.7	269	15	AAK46317	PB92 serine protea	373	6	5.7	269	15	AAK46390	Subtilisin 309 Ser
301	6	5.7	269	15	AAK46318	Subtilisin 309 Ser	374	6	5.7	269	15	AAK46391	PB92 serine protea
302	6	5.7	269	15	AAK46319	PB92 serine protea	375	6	5.7	269	15	AAK46392	Subtilisin 309 Ser



522	6	5.7	269	16	AAW87479	Subtilisin N76D/S9	595	6	5.7	269	17	AAW38314	BLS309 mutant Arg1
523	6	5.7	269	16	AAW87478	Subtilisin N76D/S9	596	6	5.7	269	17	AAW38315	Subtilisin 309 (BL
524	6	5.7	269	16	AAW87477	Subtilisin N76D/S9	597	6	5.7	269	17	AAW38316	Subtilisin 309 (BL
525	6	5.7	269	16	AAW87476	Subtilisin N76D/N1	598	6	5.7	269	17	AAW38317	BLS309 mutant Ser5
526	6	5.7	269	16	AAW87475	Subtilisin N76D/I1	599	6	5.7	269	17	AAW38318	BLS309 mutant Tyr1
527	6	5.7	269	16	AAW87474	Subtilisin N76D/V1	600	6	5.7	269	17	AAW22770	BLS309 mutant Tyr1
528	6	5.7	269	16	AAW87473	Subtilisin N76D/S1	601	6	5.7	269	17	AAW22771	BLS309 mutant Tyr1
529	6	5.7	269	16	AAW87472	Subtilisin N76D/S1	602	6	5.7	269	17	AAW22755	Subtilisin 309 (BL
530	6	5.7	269	16	AAW87471	Subtilisin N76D/S9	603	6	5.7	269	17	AAW22756	Subtilisin 309 (BL
531	6	5.7	269	16	AAW87505	Subtilisin N76D/S1	604	6	5.7	269	17	AAW22757	Subtilisin 309 (BL
532	6	5.7	269	16	AAW87504	Subtilisin N76D/S1	605	6	5.7	269	17	AAW22758	Subtilisin 309 (BL
533	6	5.7	269	16	AAW87503	Subtilisin N76D/S1	606	6	5.7	269	17	AAW22759	BLS309 mutant Ser5
534	6	5.7	269	16	AAW73193	Subtilisin N76D/V1	607	6	5.7	269	17	AAW22760	BLS309 mutant Arg1
535	6	5.7	269	16	AAW87502	Subtilisin N76D/S1	608	6	5.7	269	17	AAW22761	BLS309 mutant Ser5
536	6	5.7	269	16	AAW87501	Subtilisin N76D/S1	609	6	5.7	269	17	AAW22762	BLS309 mutant Tyr1
537	6	5.7	269	16	AAW87500	Subtilisin N76D/S1	610	6	5.7	269	17	AAW22763	BLS309 mutant Arg1
538	6	5.7	269	16	AAW87499	Subtilisin N76D/S1	611	6	5.7	269	17	AAW22764	Subtilisin 309 (BL
539	6	5.7	269	16	AAW87498	Subtilisin N76D/S1	612	6	5.7	269	17	AAW22765	Subtilisin 309 (BL
540	6	5.7	269	16	AAW87497	Subtilisin N76D/S1	613	6	5.7	269	17	AAW22766	BLS309 mutant Ser5
541	6	5.7	269	16	AAW87496	Subtilisin N76D/S1	614	6	5.7	269	17	AAW22767	BLS309 mutant Tyr1
542	6	5.7	269	16	AAW87495	Subtilisin N76D/S9	615	6	5.7	269	17	AAW22768	BLS309 mutant Tyr1
543	6	5.7	269	16	AAW87494	Subtilisin N76D/S9	616	6	5.7	269	17	AAW22769	Subtilisin 309 (BL
544	6	5.7	269	16	AAW87493	Subtilisin N76D/S9	617	6	5.7	269	18	AAW34728	Bacillus lentus su
545	6	5.7	269	16	AAW87492	Subtilisin N76D/S1	618	6	5.7	269	18	AAW34729	Bacillus lentus su
546	6	5.7	269	16	AAW87491	Subtilisin N76D/S1	619	6	5.7	269	18	AAW34730	Bacillus lentus su
547	6	5.7	269	16	AAW87490	Subtilisin N76D/S1	620	6	5.7	269	18	AAW34731	Bacillus lentus su
548	6	5.7	269	16	AAW87489	Subtilisin N76D/S9	621	6	5.7	269	18	AAW34732	Bacillus lentus su
549	6	5.7	269	16	AAW87488	Subtilisin N76D/S9	622	6	5.7	269	18	AAW34733	Bacillus lentus su
550	6	5.7	269	16	AAW87487	Subtilisin N76D/S9	623	6	5.7	269	18	AAW34734	Bacillus lentus su
551	6	5.7	269	16	AAW73171	Wild type B. lentus	624	6	5.7	269	18	AAW34735	Bacillus lentus su
552	6	5.7	269	16	AAW73172	Subtilisin N76D/S9	625	6	5.7	269	18	AAW34736	Bacillus lentus su
553	6	5.7	269	16	AAW73173	Subtilisin N76D/S1	626	6	5.7	269	18	AAW34737	Bacillus lentus su
554	6	5.7	269	16	AAW73174	Subtilisin N76D/S1	627	6	5.7	269	18	AAW34738	Bacillus lentus su
555	6	5.7	269	16	AAW73175	Subtilisin N76D/S1	628	6	5.7	269	18	AAW34739	Bacillus lentus su
556	6	5.7	269	16	AAW73176	Subtilisin N76D/I1	629	6	5.7	269	18	AAW34740	Bacillus lentus su
557	6	5.7	269	16	AAW73177	Subtilisin N76D/N1	630	6	5.7	269	18	AAW34741	Bacillus lentus su
558	6	5.7	269	16	AAW73178	Subtilisin N76D/S9	631	6	5.7	269	18	AAW34742	Bacillus lentus su
559	6	5.7	269	16	AAW73179	Subtilisin N76D/S9	632	6	5.7	269	18	AAW34763	Bacillus lentus su
560	6	5.7	269	16	AAW73180	Subtilisin N76D/S9	633	6	5.7	269	18	AAW34764	Bacillus lentus su
561	6	5.7	269	16	AAW73181	Subtilisin N76D/S1	634	6	5.7	269	18	AAW34777	Bacillus lentus su
562	6	5.7	269	16	AAW73182	Subtilisin N76D/S1	635	6	5.7	269	18	AAW34694	Bacillus lentus su
563	6	5.7	269	16	AAW73183	Subtilisin N76D/S1	636	6	5.7	269	18	AAW34695	Bacillus lentus su
564	6	5.7	269	16	AAW73184	Subtilisin N76D/V1	637	6	5.7	269	18	AAW34724	Bacillus lentus su
565	6	5.7	269	16	AAW73185	Subtilisin N76D/V1	638	6	5.7	269	18	AAW34743	Bacillus lentus su
566	6	5.7	269	16	AAW73186	Subtilisin N76D/V1	639	6	5.7	269	18	AAW34744	Bacillus lentus su
567	6	5.7	269	16	AAW73187	Subtilisin N76D/I1	640	6	5.7	269	18	AAW34745	Bacillus lentus su
568	6	5.7	269	16	AAW73198	Subtilisin N76D/S1	641	6	5.7	269	18	AAW34746	Bacillus lentus su
569	6	5.7	269	16	AAW73199	Subtilisin N76D/S1	642	6	5.7	269	18	AAW34747	Bacillus lentus su
570	6	5.7	269	16	AAW73200	Subtilisin N76D/S1	643	6	5.7	269	18	AAW34748	Bacillus lentus su
571	6	5.7	269	16	AAW62007	Subtilisin N76D/S1	644	6	5.7	269	18	AAW34749	Bacillus lentus su
572	6	5.7	269	16	AAW62008	Subtilisin N76D/S1	645	6	5.7	269	18	AAW34750	Bacillus lentus su
573	6	5.7	269	16	AAW62009	Subtilisin N76D/S1	646	6	5.7	269	18	AAW34751	Bacillus lentus su
574	6	5.7	269	16	AAW62010	Subtilisin N76D/S1	647	6	5.7	269	18	AAW34752	Bacillus lentus su
575	6	5.7	269	16	AAW62011	Subtilisin N76D/S1	648	6	5.7	269	18	AAW34753	Bacillus lentus su
576	6	5.7	269	16	AAW62012	Subtilisin N76D/S1	649	6	5.7	269	18	AAW34754	Bacillus lentus su
577	6	5.7	269	16	AAW62013	Subtilisin N76D/S1	650	6	5.7	269	18	AAW34710	Bacillus lentus su
578	6	5.7	269	16	AAW73188	Subtilisin N76D/S9	651	6	5.7	269	18	AAW34711	Bacillus lentus su
579	6	5.7	269	16	AAW73189	Subtilisin N76D/S9	652	6	5.7	269	18	AAW34712	Bacillus lentus su
580	6	5.7	269	16	AAW73190	Subtilisin N76D/S9	653	6	5.7	269	18	AAW34713	Bacillus lentus su
581	6	5.7	269	16	AAW73191	Subtilisin N76D/S1	654	6	5.7	269	18	AAW34714	Bacillus lentus su
582	6	5.7	269	16	AAW73192	Subtilisin N76D/S1	655	6	5.7	269	18	AAW34715	Bacillus lentus su
583	6	5.7	269	16	AAW73194	Subtilisin N76D/S9	656	6	5.7	269	18	AAW34716	Bacillus lentus su
584	6	5.7	269	16	AAW73195	Subtilisin N76D/S9	657	6	5.7	269	18	AAW34717	Bacillus lentus su
585	6	5.7	269	16	AAW73196	Subtilisin N76D/S9	658	6	5.7	269	18	AAW34718	Bacillus lentus su
586	6	5.7	269	16	AAW73197	Subtilisin N76D/S1	659	6	5.7	269	18	AAW34719	Bacillus lentus su
587	6	5.7	269	16	AAW38306	Subtilisin 309 (BL	660	6	5.7	269	18	AAW34720	Bacillus lentus su
588	6	5.7	269	17	AAW38307	Subtilisin 309 (BL	661	6	5.7	269	18	AAW34721	Bacillus lentus su
589	6	5.7	269	17	AAW38308	Subtilisin 309 (BL	662	6	5.7	269	18	AAW34722	Bacillus lentus su
590	6	5.7	269	17	AAW38309	Subtilisin 309 (BL	663	6	5.7	269	18	AAW34723	Bacillus lentus su
591	6	5.7	269	17	AAW38310	Subtilisin 309 (BL	664	6	5.7	269	18	AAW34725	Bacillus lentus su
592	6	5.7	269	17	AAW38311	BLS309 mutant Ser5	665	6	5.7	269	18	AAW34726	Bacillus lentus su
593	6	5.7	269	17	AAW38312	BLS309 mutant Ser5	666	6	5.7	269	18	AAW34696	Bacillus lentus su
594	6	5.7	269	17	AAW38313	BLS309 mutant Tyr1	667	6	5.7	269	18	AAW34697	Bacillus lentus su

668	6	5.7	269	18	AAW34698	Bacillus lentus su
669	6	5.7	269	18	AAW34699	Bacillus lentus su
670	6	5.7	269	18	AAW34700	Bacillus lentus su
671	6	5.7	269	18	AAW34701	Bacillus lentus su
672	6	5.7	269	18	AAW34702	Bacillus lentus su
673	6	5.7	269	18	AAW34703	Bacillus lentus su
674	6	5.7	269	18	AAW34704	Bacillus lentus su
675	6	5.7	269	18	AAW34705	Bacillus lentus su
676	6	5.7	269	18	AAW34706	Bacillus lentus su
677	6	5.7	269	18	AAW34707	Bacillus lentus su
678	6	5.7	269	18	AAW34708	Bacillus lentus su
679	6	5.7	269	18	AAW34709	Bacillus lentus su
680	6	5.7	269	18	AAW34938	Bacillus lentus su
681	6	5.7	269	18	AAW34939	Bacillus lentus su
682	6	5.7	269	18	AAW34940	Bacillus lentus su
683	6	5.7	269	18	AAW34941	Bacillus lentus su
684	6	5.7	269	18	AAW34942	Bacillus lentus su
685	6	5.7	269	18	AAW34943	Bacillus lentus su
686	6	5.7	269	18	AAW34944	Bacillus lentus su
687	6	5.7	269	18	AAW34945	Bacillus lentus su
688	6	5.7	269	18	AAW34946	Bacillus lentus su
689	6	5.7	269	18	AAW34947	Bacillus lentus su
690	6	5.7	269	18	AAW34948	Bacillus lentus su
691	6	5.7	269	18	AAW34949	Bacillus lentus su
692	6	5.7	269	18	AAW34950	Bacillus lentus su
693	6	5.7	269	18	AAW34951	Bacillus lentus su
694	6	5.7	269	18	AAW34952	Bacillus lentus su
695	6	5.7	269	18	AAW34953	Bacillus lentus su
696	6	5.7	269	18	AAW34962	Bacillus lentus su
697	6	5.7	269	18	AAW34963	Bacillus lentus su
698	6	5.7	269	18	AAW34922	Bacillus lentus su
699	6	5.7	269	18	AAW34923	Bacillus lentus su
700	6	5.7	269	18	AAW34924	Bacillus lentus su
701	6	5.7	269	18	AAW34925	Bacillus lentus su
702	6	5.7	269	18	AAW34927	Bacillus lentus su
703	6	5.7	269	18	AAW34928	Bacillus lentus su
704	6	5.7	269	18	AAW34929	Bacillus lentus su
705	6	5.7	269	18	AAW34930	Bacillus lentus su
706	6	5.7	269	18	AAW34931	Bacillus lentus su
707	6	5.7	269	18	AAW34932	Bacillus lentus su
708	6	5.7	269	18	AAW34933	Bacillus lentus su
709	6	5.7	269	18	AAW34934	Bacillus lentus su
710	6	5.7	269	18	AAW34935	Bacillus lentus su
711	6	5.7	269	18	AAW34936	Bacillus lentus su
712	6	5.7	269	18	AAW34937	Bacillus lentus su
713	6	5.7	269	18	AAW34906	Bacillus lentus su
714	6	5.7	269	18	AAW34907	Bacillus lentus su
715	6	5.7	269	18	AAW34908	Bacillus lentus su
716	6	5.7	269	18	AAW34909	Bacillus lentus su
717	6	5.7	269	18	AAW34910	Bacillus lentus su
718	6	5.7	269	18	AAW34911	Bacillus lentus su
719	6	5.7	269	18	AAW34912	Bacillus lentus su
720	6	5.7	269	18	AAW34913	Bacillus lentus su
721	6	5.7	269	18	AAW34914	Bacillus lentus su
722	6	5.7	269	18	AAW34915	Bacillus lentus su
723	6	5.7	269	18	AAW34916	Bacillus lentus su
724	6	5.7	269	18	AAW34917	Bacillus lentus su
725	6	5.7	269	18	AAW34918	Bacillus lentus su
726	6	5.7	269	18	AAW34919	Bacillus lentus su
727	6	5.7	269	18	AAW34920	Bacillus lentus su
728	6	5.7	269	18	AAW34921	Bacillus lentus su
729	6	5.7	269	18	AAW34890	Bacillus lentus su
730	6	5.7	269	18	AAW34893	Bacillus lentus su
731	6	5.7	269	18	AAW34894	Bacillus lentus su
732	6	5.7	269	18	AAW34895	Bacillus lentus su
733	6	5.7	269	18	AAW34896	Bacillus lentus su
734	6	5.7	269	18	AAW34897	Bacillus lentus su
735	6	5.7	269	18	AAW34898	Bacillus lentus su
736	6	5.7	269	18	AAW34899	Bacillus lentus su
737	6	5.7	269	18	AAW34900	Bacillus lentus su
738	6	5.7	269	18	AAW34901	Bacillus lentus su
739	6	5.7	269	18	AAW34902	Bacillus lentus su
740	6	5.7	269	18	AAW34903	Bacillus lentus su

741	6	5.7	269	18	AAW34904	Bacillus lentus su
742	6	5.7	269	18	AAW34905	Bacillus lentus su
743	6	5.7	269	18	AAW18709	Mutant BLAP F46.
744	6	5.7	269	18	AAW18710	Mutant BLAP F47.
745	6	5.7	269	18	AAW18711	Mutant BLAP F49.
746	6	5.7	269	18	AAW18712	Mutant BLAP F54.
747	6	5.7	269	18	AAW18713	Mutant BLAP F55.
748	6	5.7	269	18	AAW18705	Mutant BLAP F11.
749	6	5.7	269	18	AAW18706	Mutant BLAP F43.
750	6	5.7	269	18	AAW18707	Mutant BLAP F44.
751	6	5.7	269	18	AAW18708	Mutant BLAP F45.
752	6	5.7	269	19	AAW5706	Proteolytic enzyme
753	6	5.7	269	19	AAW62224	Subtilase BAPB92 f
754	6	5.7	269	20	AAW121651	Subtilase BUSAV1.
755	6	5.7	269	20	AAW121648	B. lentus subtilis
756	6	5.7	269	20	AAW08313	B. lentus subtilis
757	6	5.7	269	20	AAW08318	B. lentus subtilis
758	6	5.7	269	20	AAW01586	Amino acid sequenc
759	6	5.7	269	20	AAW01587	Amino acid sequenc
760	6	5.7	269	20	AAW01588	B. lentus subtilis
761	6	5.7	269	20	AAW82796	Mature wild-type s
762	6	5.7	269	21	AAW11906	Subtilisin protein
763	6	5.7	269	21	AAW03775	Bacillus lentus wi
764	6	5.7	269	21	AAW02976	Bacillus lentus mu
765	6	5.7	269	21	AAW02977	Bacillus lentus mu
766	6	5.7	269	21	AAW02978	Bacillus lentus mu
767	6	5.7	269	21	AAW02979	Bacillus lentus mu
768	6	5.7	269	21	AAW02980	Bacillus lentus mu
769	6	5.7	269	21	AAW02981	Bacillus lentus mu
770	6	5.7	269	21	AAW02982	Bacillus lentus mu
771	6	5.7	269	21	AAW02983	Bacillus lentus mu
772	6	5.7	269	21	AAW02984	Bacillus lentus mu
773	6	5.7	269	21	AAW02985	Bacillus lentus mu
774	6	5.7	269	21	AAW02986	Bacillus lentus mu
775	6	5.7	269	21	AAW02987	Bacillus lentus mu
776	6	5.7	269	21	AAW02988	Bacillus lentus mu
777	6	5.7	269	21	AAW02989	Bacillus lentus mu
778	6	5.7	269	21	AAW02990	Bacillus lentus mu
779	6	5.7	269	21	AAW02991	Bacillus lentus mu
780	6	5.7	269	21	AAW02992	Bacillus lentus mu
781	6	5.7	269	21	AAW02993	Bacillus lentus mu
782	6	5.7	269	21	AAW02994	Bacillus lentus mu
783	6	5.7	269	21	AAW02995	Bacillus lentus mu
784	6	5.7	269	21	AAW02996	Bacillus lentus mu
785	6	5.7	269	21	AAW02997	Bacillus lentus mu
786	6	5.7	269	21	AAW02998	Bacillus lentus mu
787	6	5.7	269	21	AAW02999	Bacillus lentus pr
788	6	5.7	269	21	AAW97805	Bacillus lentus Sa
789	6	5.7	269	21	AAW97806	Bacillus lentus Sa
790	6	5.7	269	21	AAW97807	Bacillus lentus Sa
791	6	5.7	269	21	AAW97808	Bacillus lentus Sa
792	6	5.7	269	21	AAW97809	Bacillus lentus Sa
793	6	5.7	269	21	AAW97810	Bacillus lentus Sa
794	6	5.7	269	21	AAW97811	Bacillus lentus Sa
795	6	5.7	269	21	AAW77006	Bacillus lentus ma
796	6	5.7	269	21	AAW77007	Bacillus lentus ma
797	6	5.7	269	21	AAW77008	Bacillus lentus ma
798	6	5.7	269	21	AAW77009	Bacillus lentus ma
799	6	5.7	269	21	AAW77010	Bacillus lentus ma
800	6	5.7	269	21	AAW77011	Bacillus lentus ma
801	6	5.7	269	21	AAW77012	Bacillus lentus ma
802	6	5.7	269	21	AAW77013	Bacillus lentus ma
803	6	5.7	269	21	AAW77014	Bacillus lentus ma
804	6	5.7	269	21	AAW77015	Bacillus lentus ma
805	6	5.7	269	21	AAW77016	Bacillus lentus ma
806	6	5.7	269	21	AAW77017	Bacillus lentus ma
807	6	5.7	269	21	AAW77018	Bacillus lentus ma
808	6	5.7	269	21	AAW77019	Bacillus lentus ma
809	6	5.7	269	21	AAW77020	Bacillus lentus ma
810	6	5.7	269	21	AAW77021	Bacillus lentus ma
811	6	5.7	269	21	AAW77022	Bacillus lentus ma
812	6	5.7	269	21	AAW77023	Bacillus lentus ma
813	6	5.7	269	21	AAW77024	Bacillus lentus ma





960	6	5.7	269	23	ABB75200	Bacillus lentus su
961	6	5.7	269	23	ABB75201	Bacillus lentus su
962	6	5.7	269	23	ABB75202	Bacillus lentus su
963	6	5.7	269	23	ABB75203	Bacillus lentus su
964	6	5.7	269	23	ABB75204	Bacillus lentus su
965	6	5.7	269	23	ABB75205	Bacillus lentus su
966	6	5.7	269	23	ABB75206	Bacillus lentus su
967	6	5.7	269	23	ABB75207	Bacillus lentus su
968	6	5.7	269	23	ABB75208	Bacillus lentus su
969	6	5.7	269	23	ABB75209	Bacillus lentus su
970	6	5.7	269	23	ABB75210	Bacillus lentus su
971	6	5.7	269	23	ABB75211	Bacillus lentus su
972	6	5.7	269	23	ABB75212	Bacillus lentus su
973	6	5.7	269	23	ABB75213	Bacillus lentus su
974	6	5.7	269	23	ABB75214	Bacillus lentus su
975	6	5.7	269	23	ABB75215	Bacillus lentus su
976	6	5.7	269	23	ABB75216	Bacillus lentus su
977	6	5.7	269	23	ABB75217	Bacillus lentus su
978	6	5.7	269	23	ABB75218	Bacillus lentus su
979	6	5.7	269	23	ABB75219	Bacillus lentus su
980	6	5.7	269	23	ABB75220	Bacillus lentus su
981	6	5.7	269	23	ABB75221	Bacillus lentus su
982	6	5.7	269	23	ABB75222	Bacillus lentus su
983	6	5.7	269	23	ABB75223	Bacillus lentus su
984	6	5.7	269	23	ABB75224	Bacillus lentus su
985	6	5.7	269	23	ABB75225	Bacillus lentus su
986	6	5.7	269	23	ABB75226	Bacillus lentus su
987	6	5.7	269	23	ABB75227	Bacillus lentus su
988	6	5.7	269	23	ABB75228	Bacillus lentus su
989	6	5.7	269	23	ABB75229	Bacillus lentus su
990	6	5.7	269	23	ABB75230	Bacillus lentus su
991	6	5.7	269	23	ABB75231	Bacillus lentus su
992	6	5.7	269	23	ABB75232	Bacillus lentus su
993	6	5.7	269	23	ABB75233	Bacillus lentus su
994	6	5.7	269	23	ABB75234	Bacillus lentus su
995	6	5.7	269	23	ABB75235	Bacillus lentus su
996	6	5.7	269	23	ABB75236	Bacillus lentus su
997	6	5.7	269	23	ABB75237	Bacillus lentus su
998	6	5.7	269	23	ABB75238	Bacillus lentus su
999	6	5.7	269	23	ABB75239	Bacillus lentus su
1000	6	5.7	269	23	ABB75240	Bacillus lentus su

## ALIGNMENTS

RESULT 1  
ID AAY67587 standard; Protein; 106 AA.  
XX AAY67587;  
AC  
XX  
DT 19-JUN-2000 (first entry)  
XX  
DE S. mansoni cercarial elastase protein fragment.  
XX  
KW Schistosoma parasite; fusion protein; cercarial elastase; vaccine;  
KM antibody response; schistosomicide.  
XX  
OS Schistosoma mansoni.  
XX  
PN EP992582-A2.  
PD 12-APR-2000.  
XX  
PF 05-OCT-1999; 99EP-0307832.  
XX  
PR 07-OCT-1998; 98GB-0021821.  
XX  
PA (UWVA-) UNIV WALES BANGOR.  
XX  
PI Doenhoff M, Sayers J;  
XX

DR WPI; 2000-259136/23.  
DR N-PSDB; AA290608.  
XX  
PT New vaccine for treatment of Schistosoma infections contains a  
PT recombinant fusion protein comprising cercarial elastase sequence fused  
PT to bacterial, phage or viral protein -  
XX  
PS Claim 3; Page 17; 26pp; English.  
XX  
CC The invention provides a vaccine comprising a recombinant fusion protein  
CC capable of eliciting immunity against Schistosoma parasites. The fusion  
CC protein comprises the 27 or 28 kDa cercarial elastase sequence of  
CC S. mansoni or an active fragment, homologue or variant, fused to a  
CC bacterial, phage or viral protein. The vaccine containing the fusion  
CC protein has been found to induce a significantly increased antibody  
CC response against schistosoma infections, compared to the use of S.  
CC mansoni cercarial elastase in its native form. The present sequence  
CC represents the S. mansoni cercarial elastase protein fragment encoded by  
CC exon 2 of the cercarial elastase gene, used in the fusion protein of the  
CC invention.  
XX  
SQ Sequence 106 AA;

Query Match 100.0%; Score 106; DB 21; length 106;  
Best Local Similarity 100.0%; Pred. No. 3, 9e-103; Indels 0; Gaps 0;  
Matches 106; Conservative 0; Mismatches 0;

QY 1 VSFLRLRNGDQGIHHQSGVAVAPGYMPSGMSARQRPPIAQTLSGFDIAIYMLAQMNL 60  
DB 1 VSFLRLRNGDQGIHHQSGVAVAPGYMPSGMSARQRPPIAQTLSGFDIAIYMLAQMNL 60  
QY 61 QSGIRVLSLPQSPDIPPGTGVFIYVGRDNDPDRKNGGILKK 106  
DB 61 QSGIRVLSLPQSPDIPPGTGVFIYVGRDNDPDRKNGGILKK 106

RESULT 2  
ID AAY67588 standard; Protein; 16 AA.  
XX AAY67588;  
AC  
XX  
DT 19-JUN-2000 (first entry)  
XX  
DE S. mansoni cercarial elastase protein fragment.  
XX  
KW Schistosoma parasite; fusion protein; cercarial elastase; vaccine;  
KM antibody response; schistosomicide.  
XX  
OS Schistosoma mansoni.  
XX  
PN EP992582-A2.  
PD 12-APR-2000.  
XX  
PF 05-OCT-1999; 99EP-0307832.  
XX  
PR 07-OCT-1998; 98GB-0021821.  
XX  
PA (UWVA-) UNIV WALES BANGOR.  
XX  
PI Doenhoff M, Sayers J;  
XX  
DR WPI; 2000-259136/23.  
DR N-PSDB; AA290608.  
XX  
PT New vaccine for treatment of Schistosoma infections contains a  
PT recombinant fusion protein comprising cercarial elastase sequence fused  
PT to bacterial, phage or viral protein -  
XX  
PS Claim 5; Page 18; 26pp; English.  
XX  
CC The invention provides a vaccine comprising a recombinant fusion protein

CC capable of eliciting immunity against Schistosoma parasites. The fusion  
 CC protein comprises the 27 or 28 kDa cercarial elastase sequence of  
 CC S. mansoni or an active fragment, homologue or variant, fused to a  
 CC bacterial, phage or viral protein. The vaccine containing the fusion  
 CC protein has been found to induce a significantly increased antibody  
 CC response against Schistosoma infections, compared to the use of S.  
 CC mansoni cercarial elastase in its native form. The present sequence  
 CC represents the S. mansoni cercarial elastase protein fragment encoded by  
 CC exon 1 of the cercarial elastase gene, used in the fusion protein of the  
 CC invention.

XX  
 SQ Sequence 16 AA;

Query Match 15.1%; Score 16; DB 21; Length 16;  
 Best Local Similarity 100.0%; Pred. No. 1.4e-09;  
 Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 85 VGYGRDNDNDPDSKKN 100  
 |||||  
 Db 1 VGYGRDNDNDPDSKKN 16

RESULT 3

AAW21482  
 ID AAW21482 standard; peptide; 10 AA.

XX  
 AC AAW21482;

XX  
 DT 30-JUL-1997 (first entry)

DE Schistosoma elastase precursor derived signal oligopeptide #2.

XX  
 KW Hydrophilic; signal oligopeptide; hydrophilicity maxima; vaccine; HIV;  
 KW competitive inhibitor; feedback regulator; synthetase; gastrin precursor;  
 KW charge; polarity; farnesyl synthetase; plasminogen activator inhibitor 1;  
 KW hydroxymethylglutaryl coenzyme A reductase; glucagon precursor; rhesus;  
 KW gonadolibetin precursor; plasminogen activator inhibitor 2; prorenin;  
 KW Alzheimer amyloid A4; corticotropin releasing factor binding protein;  
 KW apolipoprotein B; herpes virus 1 glycoprotein B; HSV1; human; OMVVS;  
 KW herpes virus 2 glycoprotein B; HSV2; collagenase; apolipoprotein A;  
 KW Treponema pallidum membrane protein; TMPA; Islet amyloid polypeptide;  
 KW fibroblast MMP1; schistosoma elastase precursor; schistosomin;  
 KW hepatitis delta antigen; rev protein; HIV; VILV; angiotensinogen.

XX Schistosoma sp.

OS  
 PN W09519568-A1.

XX  
 PD 20-JUL-1995.

XX  
 PF 12-JAN-1995; 95WO-US00575.

XX  
 PR 14-JAN-1994; 94US-0182248.

XX  
 PA (RATH/) RATH M.

XX  
 PI Rath M;

DR  
 XX WPI; 1995-263953/34.

PT Identifying signal oligopeptide(s) in protein sequence(s) - shown as  
 PT regions of max. hydrophilicity, used in modulating communication  
 PT between protein(s)

XX  
 PS Claim 5; Page 69; 88pp; English.

XX  
 CC The sequences given in AAW21201-560 represent hydrophilic signal oligo-  
 CC peptides. These signal oligopeptides are localised on the surface  
 CC of the protein and are represented by the hydrophilicity maxima of  
 CC the protein. These peptides are enriched in charged amino acids  
 CC arranged with neutral spacer amino acids. The specific signal  
 CC character of these oligopeptides is determined by a characteristic  
 CC combination of conformation and charge within the signal sequence.

CC These oligopeptides may be used as vaccines in the treatment of  
 CC human disease, as competitive inhibitors to prevent or reduce the  
 CC metabolic action or interaction of a selected protein by blocking  
 CC its specific signal sequences, or as therapeutic agents to function  
 CC as feedback regulators to reduce synthesis rate of a selected protein.  
 CC These peptides may be modified by omitting one or more amino acids at  
 CC the N- and/or C-terminal, by substituting one or more amino acids  
 CC without consideration of charge and polarity, by substituting one or  
 CC more amino acids with amino acid residues with similar charge and/or  
 CC polarity, by omitting one or more amino acids or a combination of these.

XX  
 SQ Sequence 10 AA;

Query Match 9.4%; Score 10; DB 16; Length 10;  
 Best Local Similarity 100.0%; Pred. No. 0.0018;  
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 7 RNDGQGGIHH 16  
 |||||  
 Db 1 RNDGQGGIHH 10

RESULT 4

AAW21484  
 ID AAW21484 standard; peptide; 11 AA.

XX  
 AC AAW21484;

XX  
 DT 30-JUL-1997 (first entry)

DE Schistosoma elastase precursor derived signal oligopeptide #4.

XX  
 KW Hydrophilic; signal oligopeptide; hydrophilicity maxima; vaccine; HIV;  
 KW competitive inhibitor; feedback regulator; synthetase; gastrin precursor;  
 KW charge; polarity; farnesyl synthetase; plasminogen activator inhibitor 1;  
 KW hydroxymethylglutaryl coenzyme A reductase; glucagon precursor; rhesus;  
 KW gonadolibetin precursor; plasminogen activator inhibitor 2; prorenin;  
 KW Alzheimer amyloid A4; corticotropin releasing factor binding protein;  
 KW apolipoprotein B; herpes virus 1 glycoprotein B; HSV1; human; OMVVS;  
 KW herpes virus 2 glycoprotein B; HSV2; collagenase; apolipoprotein A;  
 KW Treponema pallidum membrane protein; TMPA; Islet amyloid polypeptide;  
 KW fibroblast MMP1; schistosoma elastase precursor; schistosomin;  
 KW hepatitis delta antigen; rev protein; HIV; VILV; angiotensinogen.

XX Schistosoma sp.

OS  
 PN W09519568-A1.

XX  
 PD 20-JUL-1995.

XX  
 PF 12-JAN-1995; 95WO-US00575.

XX  
 PR 14-JAN-1994; 94US-0182248.

XX  
 PA (RATH/) RATH M.

XX  
 PI Rath M;

DR  
 XX WPI; 1995-263953/34.

PT Identifying signal oligopeptide(s) in protein sequence(s) - shown as  
 PT regions of max. hydrophilicity, used in modulating communication  
 PT between protein(s)

XX  
 PS Claim 5; Page 70; 88pp; English.

XX  
 CC The sequences given in AAW21201-560 represent hydrophilic signal oligo-  
 CC peptides. These signal oligopeptides are localised on the surface  
 CC of the protein and are represented by the hydrophilicity maxima of  
 CC the protein. These peptides are enriched in charged amino acids  
 CC arranged with neutral spacer amino acids. The specific signal  
 CC character of these oligopeptides is determined by a characteristic  
 CC combination of conformation and charge within the signal sequence.





Db 62 VKVAFGY 68

RESULT 7  
AAR8326

ID AAR8326 standard; Protein; 341 AA.

AC AAR8326;

DT 07-AUG-1996 (first entry)

DE Human alpha 2,8-sialyltransferase.

KW Human melanoma; SK-Mel-28; ganglioside 3-synthase; tumour antigen;

KW detection; Northern blotting; glycosylation.

OS Homo sapiens.

FT Key Location/Qualifiers

FT Modified-site 56 /label= N-glycosylation\_site

FT Region 120..168 /label= sialyl\_motif

FT Modified-site 230 /label= N-glycosylation\_site

FT Region 258..280 /label= sialyl\_motif

FT JPO7327678-A.

PD 19-DEC-1995.

PF 03-JUN-1994; 94JP-0145450.

PR 03-JUN-1994; 94JP-0145450.

PA (SEBK ) SEIKAGAKU KOGYO CO LTD.

DR WPI; 1996-072336/08.

DR N-PSDB; AAT03694.

PT Alpha 2,8-sialyltransferase cDNA - useful for sensitive detection

PS of tumour antigen by Northern blotting

PS Claim 1; Page 6-7; 10pp; Japanese.

CC A cDNA clone coding for alpha 2,8-sialyltransferase was isolated

CC from a human melanoma SK-Mel-28 cDNA library and sequenced. The

CC insert codes for a protein (i.e. the present sequence) having

CC ganglioside 3-synthase activity which is useful as a biochemical

CC reagent. The cDNA is useful for detecting tumour antigen by Northern

CC blot analysis.

SQ Sequence 341 AA;

QY 100 NGGILKK 106

DB 128 NGGILKK 134

RESULT 8  
AAR63225

AC AAR63225;

DT 26-JUN-1995 (first entry)

DE Human alpha-2,8-sialyltransferase protein.

KW Human; alpha-2,8-sialyltransferase; sialyltransferase; melanoma; E.coli;

KW pUC119; expression vector; pAMOPRC3sc; PCR; amplify; primer; sugar chain;

KW physiological; ganglioside.

OS Homo sapiens.

PN WC9423020-A.

PD 13-OCT-1994.

PF 28-MAR-1994; 94WO-JP00495.

PR 29-MAR-1993; 93JP-0069988.

PA (KYOW ) KYOWA HAKKO KOGYO KK.

PI Hanai N, Kurata K, Nishi T, Sasaki K;

DR WPI; 1994-333182/41.

DR N-PSDB; AAQ77831.

PT New alpha-2,8-sialyl transferase - useful for producing

PS physiologically active sugar chains

PS Claim 1; Page 63-66; 102pp; Japanese.

CC The amino acid sequence of the novel human alpha-2,8-sialyltransferase.

CC The gene encodes a protein of 356 a.a. The gene (called WPI) was

CC isolated from the human melanoma cell line WM266-4 and cloned into the

CC E.coli plasmid pUC119 to create pUC119.WPI. The gene was inserted into

CC the expression vector pAMOPRC3sc to produce the plasmid pAMOPRC3sc.WPI.

CC The missing N-terminal 19 a.a. were inserted into this vector by PCR

CC amplification using the primers AAQ77835-6 Alpha-2,8-sialyltransferase

CC is useful in the production of physiologically active sugar chains

CC e.g. in the conversion of ganglioside GM3 to GD3.

SQ Sequence 356 AA;

QY 100 NGGILKK 106

DB 143 NGGILKK 149

RESULT 9  
AAB86207

ID AAB86207 standard; protein; 356 AA.

AC AAB86207;

DT 28-AUG-2001 (first entry)

DE Human GD3 synthase protein.

DE GD3 synthase; human; inhibitor; ganglioside; neurological disorder;

KW anti-ischemic; cerebral; neurological; antineurodegeneration;

KW antitumor activity; veterinary medicine; cerebral ischemia; stroke;

KW traumatic brain injury; spinal cord injury; neurodegenerative disease;

KW Alzheimer's disease; Parkinson's disease; amyotrophic lateral sclerosis;

KW multiple sclerosis; brain tumor.

OS Homo sapiens.

PN DE19958684-A1.

PD 07-JUN-2001.

PF 06-DEC-1999; 99DE-1058684.

XX 06-DEC-1999; 99DE-1058684.  
 XX (KNOL ) KNOL AG.  
 XX  
 XX Martin-Villaalba A, Schenkel J, Kleber S, Testi R;  
 XX WPI; 2001-357289/38.  
 XX  
 XX Use of ganglioside GD3 synthase inhibitors for treating neurological  
 XX diseases, e.g. cerebral ischemia or neurodegeneration -  
 XX  
 XX Disclosure; Page 10-11; 16pp; German.  
 XX  
 XX This invention describes a novel use of inhibitors (I) of ganglioside GD3  
 XX synthase (II) for treatment of neurological disorders and their  
 XX associated indications, symptoms and dysfunctions. The products of the  
 XX invention have anti-ischemic, cerebral, neurological,  
 XX antineurodegeneration and antitumor activity. (I) are used, in human or  
 XX veterinary medicine, to treat cerebral ischemia, (acute) stroke,  
 XX traumatic brain or spinal cord injuries, and neurodegenerative diseases  
 XX (especially Alzheimer's and Parkinson's diseases, amyotrophic lateral  
 XX sclerosis and multiple sclerosis), also brain tumors that cause injury to  
 XX the brain. This sequence represents the human GD3 synthase described in  
 XX the method of the invention.  
 XX  
 XX Sequence 356 AA;  
 SQ

Query Match 6.6%; Score 7; DB 22; Length 356;  
 Best Local Similarity 100.0%; Pred. No. 68;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 100 NGGILKK 106  
 |||||  
 Db 143 NGGILKK 149

RESULT 10  
 AAU11996  
 ID AAU11996 standard; Protein; 359 AA.  
 XX  
 AC AAU11996;  
 XX  
 XX 09-APR-2002 (first entry)  
 DT  
 XX  
 DE Klebsiella pneumoniae beta-lactamase-like (cephalosporinase) protein.  
 XX  
 KM Polymer crossover location; crossover recombination; hybrid biopolymer;  
 KM directed evolution of polymer; beta-lactamase-like protein;  
 KM cephalosporinase.  
 XX  
 OS Klebsiella pneumoniae.  
 XX  
 PN WO200190346-A2.  
 XX  
 PD 29-NOV-2001.  
 XX  
 PF 23-MAY-2001; 2001WO-US16831.  
 XX  
 XX 23-MAY-2000; 2000US-207048P.  
 PR 27-SEP-2000; 2000US-235960P.  
 PR 13-APR-2001; 2001US-283567P.  
 XX  
 PA (CALY ) CALIFORNIA INST OF TECHNOLOGY.  
 XX  
 PI Wang Z, Voigt CA, Mayo SL, Arnold FH;  
 XX  
 DR WPI; 2002-122019/16.  
 XX  
 XX  
 XX Selecting a crossover locations in biopolymers, useful for particularly  
 XX useful designing, engineering and generating new proteins and genes  
 XX with useful properties, by determining the crossover disruption  
 XX profiles of biopolymers -

XX Examples; Fig 3; 139pp; English.  
 PS  
 XX  
 XX The present invention relates to methods for identifying crossover  
 CC locations in a polymer including nucleic acids and proteins. The  
 CC method involves selecting a crossover location in a first biopolymer  
 CC having a first polymer sequence, for recombination with one or more  
 CC second biopolymers each having its own polymer sequence. The method  
 CC comprises determining a crossover disruption related to the number of  
 CC coupling interactions disrupted in the crossover mutant represented  
 CC by a data structure. The method is useful for the directed evolution  
 CC of polymers, including directed evolution of nucleic acids and proteins.  
 CC The methods are particularly useful for designing, engineering and  
 CC generating hybrid biopolymers. The methods are also useful for  
 CC accelerating the production of new proteins and genes with novel and  
 CC useful properties. The present sequence representing Klebsiella  
 CC pneumoniae beta-lactamase-like (cephalosporinase) protein is used in  
 CC crossover studies in the methods of the present invention.  
 CC  
 XX Sequence 359 AA;  
 SQ

Query Match 6.6%; Score 7; DB 23; Length 359;  
 Best Local Similarity 100.0%; Pred. No. 69;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 49 IAIWMLA 55  
 |||||  
 Db 332 IAIWMLA 338

RESULT 11  
 ABG04674  
 ID ABG04674 standard; Protein; 402 AA.  
 XX  
 AC ABG04674;  
 XX  
 XX 13-FEB-2002 (first entry)  
 DT  
 XX  
 DE Novel human diagnostic protein #465.  
 XX  
 KM Human; chromosome mapping; gene mapping; gene therapy; forensic;  
 KM food supplement; medical imaging; diagnostic; genetic disorder.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO200175067-A2.  
 XX  
 PD 11-OCT-2001.  
 XX  
 PF 30-MAR-2001; 2001WO-US08631.  
 XX  
 PR 31-MAR-2000; 2000US-0540217.  
 PR 23-AUG-2000; 2000US-0649167.  
 XX  
 PA (HYSE-) HYSEQ INC.  
 XX  
 PI Drmanac RT, Liu C, Tang YT;  
 XX  
 DR WPI; 2001-639362/73.  
 DR N-PSDB; AAS68861.  
 XX  
 XX New isolated polynucleotide and encoded polypeptides, useful in  
 PT diagnostics, forensics, gene mapping, identification of mutations  
 PT responsible for genetic disorders or other traits and to assess  
 PT biodiversity -  
 XX  
 PS Claim 20; SEQ ID No 35033; 103pp; English.  
 XX  
 XX The invention relates to isolated polynucleotide (I) and  
 CC polypeptide (II) sequences. (I) is useful as hybridisation probes,  
 CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome  
 CC and gene mapping, and in recombinant production of (II). The  
 CC polynucleotides are also used in diagnostics as expressed sequence tags

CC for identifying expressed genes. (I) is useful in gene therapy techniques  
CC to restore normal activity of (II) or to treat disease states involving  
CC (II). (II) is useful for generating antibodies against it, detecting or  
CC quantitating a polypeptide in tissue, as molecular weight markers and as  
CC a food supplement. (II) and its binding partners are useful in medical  
CC imaging of sites expressing (II). (I) and (II) are useful for treating  
CC disorders involving aberrant protein expression or biological activity.  
CC The polypeptide and polynucleotide sequences have applications in  
CC diagnostics, forensics, gene mapping, identification of mutations  
CC responsible for genetic disorders or other traits to assess biodiversity  
CC and to produce other types of data and products dependent on DNA and  
CC amino acid sequences. ABG00010-ABG30377 represent novel human  
CC diagnostic amino acid sequences of the invention.  
CC Note: The sequence data for this patent did not appear in the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pct\_sequences.  
XX  
SQ Sequence 402 AA;  
Query Match 6.6%; Score 7; DB 22; Length 402;  
Best Local Similarity 100.0%; Pred. No. 76;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
OY 57 MVNLQSG 63  
|||  
Db 138 MVNLQSG 144  
RESULT 12  
AAB14350  
ID AAB14350 standard; Protein; 725 AA.  
XX  
AC AAB14350;  
XX  
DT 29-NOV-2000 (first entry)  
XX  
DE Human partial myelin transcription factor 1.  
XX  
KW Human: myelin transcription factor 1; MYT1;  
KW epidermal cell transdifferentiation; gene therapy; cerebroprotective;  
KW neuroprotective; brain injury; spinal cord injury; stroke;  
KW neurodegenerative disease; Parkinson's disease; Huntington's disease;  
KW Alzheimer's disease; neuronal cell generation.  
XX  
OS Homo sapiens.  
XX  
PM US6087168-A.  
XX  
PD 11-JUL-2000.  
XX  
PF 20-JAN-1999; 99US-0234332.  
XX  
PR 20-JAN-1999; 99US-0234332.  
XX  
PA (CEDA-) CEDARS SINAI MEDICAL CENT.  
XX  
PI Levesque MF, Neuman T;  
XX  
DR WPI; 2000-498200/44.  
DR N-PSDB; AAA62684.  
XX  
PT Converting epidermal cells into neurons, useful for isolating nerve  
PT growth factors or for gene therapy, comprises differentiating cells  
PT and transfecting with vectors with a CDNA coding neurogenic  
PT transcription factors -  
XX  
XX Example 2; Column 39-44; 27pp; English.  
XX  
CC The present sequence is myelin transcription factor 1 (MYT1), which  
CC is encoded by a human gene in Genbank. The Genbank sequence was used to  
CC provide sequence information for the cloning of MYT1 cDNA, which was  
CC used to transfect cultured epidermal cells. This was part of a novel  
CC method for transdifferentiating an epidermal basal cell into a cell

CC having the morphological, physiological and/or immunological features of  
CC a viable neuronal cell. The method is useful for screening new drugs for  
CC treating a nervous system disorder, or for isolating a novel nerve  
CC growth factor. The transdifferentiated cell is useful in both cell and  
CC gene therapies aimed at alleviating various neurological disorders. The  
CC cell or gene therapy approach involves the use of autologous  
CC transplantation or grafting of the newly created neuronal cells as  
CC treatment for brain or spinal cord injury, stroke and neurodegenerative  
CC diseases (e.g. Parkinson's disease, Huntington's disease or Alzheimer's  
CC disease).  
XX  
SQ Sequence 725 AA;  
Query Match 6.6%; Score 7; DB 21; Length 725;  
Best Local Similarity 100.0%; Pred. No. 1.3e+02;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
OY 19 SGVKVAP 25  
|||  
Db 478 SGVKVAP 484  
RESULT 13  
AAG48794  
ID AAG48794 standard; Protein; 810 AA.  
XX  
AC AAG48794;  
XX  
DT 18-OCT-2000 (first entry)  
XX  
DE Arabidopsis thaliana protein fragment SEQ ID NO: 61656.  
XX  
KW Protein identification; signal transduction pathway; metabolic pathway;  
KW hybridisation assay; genetic mapping; gene expression control; promoter;  
KW termination sequence.  
XX  
OS Arabidopsis thaliana.  
XX  
PN EP1033405-A2.  
XX  
PD 06-SEP-2000.  
XX  
PF 25-FEB-2000; 2000EP-0301439.  
XX  
PR 25-FEB-1999; 99US-0121825.  
PR 05-MAR-1999; 99US-0123180.  
PR 09-MAR-1999; 99US-0123548.  
PR 23-MAR-1999; 99US-0125788.  
PR 25-MAR-1999; 99US-0126264.  
PR 29-MAR-1999; 99US-0126785.  
PR 01-APR-1999; 99US-0127462.  
PR 06-APR-1999; 99US-0128234.  
PR 08-APR-1999; 99US-0128714.  
PR 16-APR-1999; 99US-0129845.  
PR 19-APR-1999; 99US-0130077.  
PR 21-APR-1999; 99US-0130449.  
PR 23-APR-1999; 99US-0130510.  
PR 23-APR-1999; 99US-0130891.  
PR 28-APR-1999; 99US-0131448.  
PR 30-APR-1999; 99US-0132048.  
PR 30-APR-1999; 99US-0132407.  
PR 04-MAY-1999; 99US-0132484.  
PR 05-MAY-1999; 99US-0132485.  
PR 06-MAY-1999; 99US-0132486.  
PR 06-MAY-1999; 99US-0132487.  
PR 07-MAY-1999; 99US-0132863.  
PR 11-MAY-1999; 99US-0134256.  
PR 14-MAY-1999; 99US-0134218.  
PR 14-MAY-1999; 99US-0134219.  
PR 14-MAY-1999; 99US-0134221.  
PR 14-MAY-1999; 99US-0134370.  
PR 18-MAY-1999; 99US-0134768.  
PR 19-MAY-1999; 99US-0134941.



PR 20-MAY-1999; 99US-0135124.  
PR 21-MAY-1999; 99US-0135353.  
PR 24-MAY-1999; 99US-0135629.  
PR 25-MAY-1999; 99US-0136021.  
PR 27-MAY-1999; 99US-0136392.  
PR 28-MAY-1999; 99US-0136782.  
PR 01-JUN-1999; 99US-0137222.  
PR 03-JUN-1999; 99US-0137528.  
PR 04-JUN-1999; 99US-0137502.  
PR 07-JUN-1999; 99US-0137724.  
PR 08-JUN-1999; 99US-0138094.  
PR 10-JUN-1999; 99US-0138540.  
PR 10-JUN-1999; 99US-0138847.  
PR 14-JUN-1999; 99US-0139119.  
PR 16-JUN-1999; 99US-0139452.  
PR 16-JUN-1999; 99US-0139453.  
PR 17-JUN-1999; 99US-0139492.  
PR 18-JUN-1999; 99US-0139454.  
PR 18-JUN-1999; 99US-0139455.  
PR 18-JUN-1999; 99US-0139456.  
PR 18-JUN-1999; 99US-0139457.  
PR 18-JUN-1999; 99US-0139458.  
PR 18-JUN-1999; 99US-0139459.  
PR 18-JUN-1999; 99US-0139460.  
PR 18-JUN-1999; 99US-0139461.  
PR 18-JUN-1999; 99US-0139462.  
PR 18-JUN-1999; 99US-0139463.  
PR 18-JUN-1999; 99US-0139750.  
PR 18-JUN-1999; 99US-0139760.  
PR 21-JUN-1999; 99US-0139817.  
PR 22-JUN-1999; 99US-0139899.  
PR 23-JUN-1999; 99US-0140353.  
PR 23-JUN-1999; 99US-0140354.  
PR 24-JUN-1999; 99US-0140698.  
PR 28-JUN-1999; 99US-0140823.  
PR 29-JUN-1999; 99US-0140991.  
PR 30-JUN-1999; 99US-0141287.  
PR 01-JUL-1999; 99US-0141842.  
PR 01-JUL-1999; 99US-0142154.  
PR 02-JUL-1999; 99US-0142055.  
PR 06-JUL-1999; 99US-0142390.  
PR 08-JUL-1999; 99US-0142803.  
PR 09-JUL-1999; 99US-0142920.  
PR 12-JUL-1999; 99US-0142977.  
PR 13-JUL-1999; 99US-0143542.  
PR 14-JUL-1999; 99US-0143624.  
PR 15-JUL-1999; 99US-0144005.  
PR 16-JUL-1999; 99US-0144085.  
PR 16-JUL-1999; 99US-0144086.  
PR 19-JUL-1999; 99US-0144325.  
PR 19-JUL-1999; 99US-0144331.  
PR 19-JUL-1999; 99US-0144332.  
PR 19-JUL-1999; 99US-0144333.  
PR 19-JUL-1999; 99US-0144334.  
PR 19-JUL-1999; 99US-0144335.  
PR 20-JUL-1999; 99US-0144352.  
PR 20-JUL-1999; 99US-0144632.  
PR 20-JUL-1999; 99US-0144884.  
PR 21-JUL-1999; 99US-0144814.  
PR 21-JUL-1999; 99US-0145086.  
PR 21-JUL-1999; 99US-0145087.  
PR 22-JUL-1999; 99US-0145085.  
PR 22-JUL-1999; 99US-0145086.  
PR 22-JUL-1999; 99US-0145087.  
PR 22-JUL-1999; 99US-0145088.  
PR 22-JUL-1999; 99US-0145192.  
PR 23-JUL-1999; 99US-0145145.  
PR 23-JUL-1999; 99US-0145218.  
PR 23-JUL-1999; 99US-0145224.  
PR 26-JUL-1999; 99US-0145275.  
PR 27-JUL-1999; 99US-0145913.  
PR 27-JUL-1999; 99US-0145918.  
PR 27-JUL-1999; 99US-0145919.  
PR 28-JUL-1999; 99US-0145951.  
  
PR 02-AUG-1999; 99US-0146386.  
PR 02-AUG-1999; 99US-0146388.  
PR 02-AUG-1999; 99US-0146389.  
PR 03-AUG-1999; 99US-0147038.  
PR 04-AUG-1999; 99US-0147204.  
PR 04-AUG-1999; 99US-0147302.  
PR 05-AUG-1999; 99US-0147192.  
PR 05-AUG-1999; 99US-0147260.  
PR 06-AUG-1999; 99US-0147303.  
PR 06-AUG-1999; 99US-0147416.  
PR 09-AUG-1999; 99US-0147493.  
PR 09-AUG-1999; 99US-0147935.  
PR 10-AUG-1999; 99US-0148171.  
PR 11-AUG-1999; 99US-0148319.  
PR 12-AUG-1999; 99US-0148341.  
PR 13-AUG-1999; 99US-0148565.  
PR 13-AUG-1999; 99US-0148684.  
PR 16-AUG-1999; 99US-0149368.  
PR 17-AUG-1999; 99US-0149175.  
PR 18-AUG-1999; 99US-0149426.  
PR 20-AUG-1999; 99US-0149722.  
PR 20-AUG-1999; 99US-0149723.  
PR 20-AUG-1999; 99US-0149929.  
PR 23-AUG-1999; 99US-0149902.  
PR 23-AUG-1999; 99US-0149930.  
PR 25-AUG-1999; 99US-0150566.  
PR 26-AUG-1999; 99US-0150884.  
PR 27-AUG-1999; 99US-0151065.  
PR 27-AUG-1999; 99US-0151066.  
PR 27-AUG-1999; 99US-0151080.  
PR 30-AUG-1999; 99US-0151303.  
PR 31-AUG-1999; 99US-0151438.  
PR 01-SEP-1999; 99US-0151930.  
PR 07-SEP-1999; 99US-0152353.  
PR 10-SEP-1999; 99US-0153070.  
PR 13-SEP-1999; 99US-0153758.  
PR 16-SEP-1999; 99US-0154018.  
PR 20-SEP-1999; 99US-0154039.  
PR 22-SEP-1999; 99US-0155139.  
PR 23-SEP-1999; 99US-0155486.  
PR 24-SEP-1999; 99US-0155659.  
PR 28-SEP-1999; 99US-0156458.  
PR 29-SEP-1999; 99US-0156596.  
PR 04-OCT-1999; 99US-0157117.  
PR 05-OCT-1999; 99US-0157753.  
PR 06-OCT-1999; 99US-0157865.  
PR 07-OCT-1999; 99US-0158029.  
PR 08-OCT-1999; 99US-0158232.  
PR 12-OCT-1999; 99US-0158369.  
PR 13-OCT-1999; 99US-0159293.  
PR 13-OCT-1999; 99US-0159294.  
PR 13-OCT-1999; 99US-0159295.  
PR 13-OCT-1999; 99US-0159329.  
PR 14-OCT-1999; 99US-0159330.  
PR 14-OCT-1999; 99US-0159331.  
PR 14-OCT-1999; 99US-0159637.  
PR 14-OCT-1999; 99US-0159638.  
PR 18-OCT-1999; 99US-0159584.  
PR 21-OCT-1999; 99US-0160741.  
PR 21-OCT-1999; 99US-0160767.  
PR 21-OCT-1999; 99US-0160768.  
PR 21-OCT-1999; 99US-0160770.  
PR 21-OCT-1999; 99US-0160770.  
PR 21-OCT-1999; 99US-0160814.  
PR 21-OCT-1999; 99US-0160815.  
PR 21-OCT-1999; 99US-0160815.  
PR 22-OCT-1999; 99US-0160980.  
PR 22-OCT-1999; 99US-0160981.  
PR 25-OCT-1999; 99US-0160989.  
PR 25-OCT-1999; 99US-0161405.  
PR 25-OCT-1999; 99US-0161405.  
PR 25-OCT-1999; 99US-0161406.  
PR 26-OCT-1999; 99US-0161359.  
PR 26-OCT-1999; 99US-0161360.

PR 26-OCT-1999; 99US-0161361.  
 PR 28-OCT-1999; 99US-0161920.  
 PR 28-OCT-1999; 99US-0161992.  
 PR 28-OCT-1999; 99US-0161993.  
 PR 29-OCT-1999; 99US-0162142.

Query Match 6.6%; Score 7; DB 21; Length 810;  
 Best Local Similarity 100.0%; Pred. No. 1.5e+02;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 21 VKVAPGY 27  
 Db 62 VKVAPGY 68

## RESULT 14

AAAB42339  
 ID AAB42339 standard; Protein; 829 AA.

AC AAB42339;

DT 08-FEB-2001 (first entry)

DE Human ORFX ORF2103 polypeptide sequence SEQ ID NO:4206.

XX Human; open reading frame; ORFX; detection; cytosolic; hepatotropic;  
 KW vulnerability; antiparkinsonian; neurotrophic; neuroprotective;  
 KW anticonvulsant; osteoporotic; antiarthritic; immunosuppressant; cardiant;  
 KW immunostimulant; thrombolytic; coagulant; vasotropic; antidiabetic;  
 KW hypotensive; dermatological; immunosuppressive; antiinflammatory;  
 KW antiviral; antibacterial; antifungal; antineumatic; antihydrolyd;  
 KW antianaemic; gene therapy; cancer; proliferative disorder; hypertension;  
 KW neurodegenerative disorder; osteoarthritis; graft vs host disease;  
 KW cardiovascular disease; diabetes mellitus; hypothyroidism; SCID; AIDS;  
 KW cholesterol ester storage; systemic lupus erythematosus; infection;  
 KW severe combined immunodeficiency; malaria; autoimmune disorder; asthma;  
 KW allergy; aplastic anaemia; nocturnal haemoglobinuria; burn; wound;  
 KW bone damage; cartilage damage; antiinflammatory disease; coagulation;  
 KW thrombosis; contraceptive.

OS Homo sapiens.

XX WO200058473-A2.

XX 05-OCT-2000.

XX 31-MAR-2000; 2000WO-US08621.

XX 31-MAR-1999; 99US-0127607.

PR 02-APR-1999; 99US-0127636.

PR 05-APR-1999; 99US-0127728.

PR 30-MAR-2000; 2000US-0540763.

XX (CDRA-) CDRAGEN CORP.

XX Shimketa RA, Leach M;

XX WPI; 2000-602362/57.

DR N-PSDB; AAC76548.

XX Novel nucleic acids and peptides derived from open reading frame X,  
 PT useful for treating e.g. cancers, proliferative disorders,  
 PT neurodegenerative disorders and cardiovascular disease -  
 PS Claim 11; Page 3402-3404; 5507pp; English.

XX AAC74446 to AAC77606 encode the proteins given in AAB40237 to AAB43397,  
 CC sequences have activities such as: cytosolic; hepatotropic; vulnary;  
 CC antiparkinsonian; neurotrophic; neuroprotective;  
 CC osteoporotic; anticonvulsant; antidiabetic; immunosuppressant;  
 CC immunostimulant; cardiant; thrombolytic; coagulant; vasotropic;  
 CC antidiabetic; hypotensive; dermatological; immunosuppressive;

CC antiinflammatory; antibacterial; antiviral; antifungal; antineumatic;  
 CC antihydrolyd; and antianaemic. The sequences can be used for determining  
 CC the presence of or predisposition to, or preventing or treating  
 CC pathological conditions associated with an ORFX-associated disorder. The  
 CC nucleic acids can be used to express ORFX proteins in gene therapy  
 CC vectors. The proteins and nucleic acids may be used to treat cancers,  
 CC proliferative disorders, neurodegenerative disorders, osteoarthritis,  
 CC graft vs host disease, cardiovascular disease, diabetes mellitus,  
 CC hypertension, hypothyroidism, cholesterol ester storage, systemic lupus  
 CC erythematosus, severe combined immunodeficiency (SCID), AIDS, viral,  
 CC bacterial or fungal infection, malaria, autoimmune disorders, asthma,  
 CC allergies, aplastic anaemia, burns, wounds, bone and cartilage damage,  
 CC nocturnal haemoglobinuria, antiinflammatory disease; to enhance  
 CC coagulation; to inhibit thrombosis; and as a contraceptive.

XX Sequence 829 AA;

Query Match 6.6%; Score 7; DB 21; Length 829;  
 Best Local Similarity 100.0%; Pred. No. 1.5e+02;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 19 SGVAVAP 25  
 Db 582 SGVAVAP 588

## RESULT 15

ABBB4668  
 ID ABB4668 standard; Protein; 838 AA.

AC ABB4668;

DT 26-MAR-2002 (first entry)

DE Drosophila melanogaster polypeptide SEQ ID NO 20796.

XX Drosophila; developmental biology; cell signalling; insecticide;  
 KW pharmacological.

XX Drosophila melanogaster.

OS Drosophila melanogaster.

XX WO200171042-A2.

XX 27-SEP-2001.

XX 23-MAR-2001; 2001WO-US09231.

XX 23-MAR-2000; 2000US-191637P.

PR 11-JUL-2000; 2000US-0614150.

PA (PEKE ) PE CORP NY.

PI Venter JC, Adams M, Li PWD, Myers EW;

XX WPI; 2001-656860/75.

DR N-PSDB; ABL08771.

XX New isolated nucleic acid detection reagent for detecting 1000 or more  
 PT genes from Drosophila and for elucidating cell signalling and cell-cell  
 PT interactions -  
 PS Disclosure; SEQ ID NO 20796; 21pp + Sequence Listing; English.

XX The invention relates to an isolated nucleic acid detection reagent  
 CC capable of detecting 1000 or more genes from Drosophila. The invention is  
 CC useful in developmental biology and in elucidating cell signalling and  
 CC cell-cell interactions in higher eukaryotes for the development of  
 CC insecticides, therapeutics and pharmaceutical drugs. The invention  
 CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA  
 CC sequences (ABL01840-ABL16175) and the encoded proteins  
 CC (ABBS7737-ABBS72072).  
 CC The sequence data for this patent did not form part of the printed  
 CC specification, but was obtained in electronic format directly from WIPO

CC at ftp.wipo.int/pub/published\_pct\_sequences.  
XX  
SQ Sequence 838 AA;  
Query Match 6.6%; Score 7; DB 22; Length 838;  
Best Local Similarity 100.0%; Pred. No. 1.5e+02;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
OY 72 PSDIPP 78  
| | | | | |  
Db 300 PSDIPP 306  
RESULT 16  
AAU08681  
ID AAU08681 standard; Protein; 2724 AA.  
XX  
AC AAU08681;  
XX  
DT 18-DEC-2001 (first entry)  
XX  
DE Human FCTR3f polypeptide sequence.  
XX  
KW Human; FCTR3; myelogenous leukaemia; carcinoma; melanoma; glioma;  
KW astrocytoma; congenital neonatal alloimmune thrombocytopenia; infection;  
KW neurological disorder; neurodegenerative disorders; nerve trauma;  
KW familial myelodysplastic syndrome; Charcot-Marie-Tooth neuropathy;  
KW demyelinating Gardner syndrome; familial myelodysplastic syndrome;  
KW mental health condition; immunological disorder; allergy; infertility;  
KW bronchial asthma; Avelino type eosinophilia; lung disease; deafness;  
KW reproductive disorder; reproductive disorder; glycoprotein Ia deficiency;  
KW desmoid disease; tuncot syndrome; liver cirrhosis; hepatitis C; virucide;  
KW gastric disorders; pancreatic disease; Schistosoma mansoni infection;  
KW spinocerebellar ataxia; Plasmodium falciparum parasitaemia; diabetes;  
KW Corneal dystrophy-Greponow type I; Corneal dystrophy-lattice type I;  
KW Reis-Bucklers corneal dystrophy; cytostatic; immunosuppressive;  
KW antiallergic; antiaesthetic; antiinfectivity; antiinflammatory;  
KW antidiabetic; protozoacide; hepatotropic; virucide; ophthalmological;  
KW gynaecological; antiinfertility; immunostimulant; auditory; haemostatic;  
KW gene therapy; FCTR3a; neurestin-like protein; FCTR3f.  
XX  
OS Homo sapiens.  
XX  
PN WO20016747-A2.  
XX  
PD 13-SEP-2001.  
XX  
PE 05-MAR-2001; 2001WO-US07160.  
XX  
PF 03-MAR-2000; 2000US-186592P.  
PR 03-MAR-2000; 2000US-186718P.  
PR 06-MAR-2000; 2000US-187293P.  
PR 06-MAR-2000; 2000US-187294P.  
PR 17-MAR-2000; 2000US-190400P.  
PR 07-APR-2000; 2000US-196018P.  
PR 03-JAN-2001; 2001US-259548P.  
XX  
PA (CURA-) CURAGEN CORP.  
XX  
PI Vernet CAM, Fernandes E, Shimkets RA, Herrmann JL, Majumder K;  
PI MacDougall J, Mishra V, Mezes PS, Rastelli L;  
XX  
DR WPI; 2001-596837/67.  
DR N-PSDB; AAS14089.  
XX  
PT Novel polypeptides designated as FCTR3 polypeptides, useful in  
PT detection, prevention and treatment of a broad range of pathological  
PT states -  
XX  
PS Claim 1; Page 39; 215pp; English.  
XX  
CC The invention relates to human FCTR3 polypeptides, FCTR1-FCTR7, and the  
CC nucleic acids encoding them. These sequences are useful for the treatment

CC or prevention of numerous disorders including myelogenous leukemia,  
CC carcinomas, melanomas, gliomas, astrocytomas, congenital neonatal  
CC alloimmune thrombocytopenia, neurological disorders, neurodegenerative  
CC disorders, nerve trauma, familial myelodysplastic syndrome,  
CC Charcot-Marie-Tooth neuropathy, demyelinating Gardner syndrome, familial  
CC myelodysplastic syndrome, mental health conditions, immunological  
CC disorders, allergy and infection, bronchial asthma, Avelino type  
CC eosinophilia, lung diseases, reproductive disorders, infertility, male  
CC and female reproductive disorders, deafness, glycoprotein Ia deficiency,  
CC desmoid disease, tuncot syndrome, liver cirrhosis, hepatitis C, gastric  
CC disorders, pancreatic diseases such as diabetes, Schistosoma mansoni  
CC infection spinocerebellar ataxia, Plasmodium falciparum parasitaemia,  
CC Corneal dystrophy-Greponow type I, Corneal dystrophy-lattice type I and  
CC Reis-Bucklers corneal dystrophy. This sequence represents FCTR3f, a  
CC homologue of FCTR3a protein.  
XX  
SQ Sequence 2724 AA;  
XX  
Query Match 6.6%; Score 7; DB 22; Length 2724;  
Best Local Similarity 100.0%; Pred. No. 4.5e+02;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
OY 57 MYNLSG 63  
| | | | | |  
Db 2001 MYNLSG 2007  
RESULT 17  
AAU08680  
ID AAU08680 standard; Protein; 2733 AA.  
XX  
AC AAU08680;  
XX  
DT 18-DEC-2001 (first entry)  
XX  
DE Human FCTR3b polypeptide sequence.  
XX  
KW Human; FCTR3; myelogenous leukaemia; carcinoma; melanoma; glioma;  
KW astrocytoma; congenital neonatal alloimmune thrombocytopenia; infection;  
KW neurological disorder; neurodegenerative disorders; nerve trauma;  
KW familial myelodysplastic syndrome; Charcot-Marie-Tooth neuropathy;  
KW demyelinating Gardner syndrome; familial myelodysplastic syndrome;  
KW mental health condition; immunological disorder; allergy; infertility;  
KW bronchial asthma; Avelino type eosinophilia; lung disease; deafness;  
KW reproductive disorder; reproductive disorder; glycoprotein Ia deficiency;  
KW desmoid disease; tuncot syndrome; liver cirrhosis; hepatitis C; virucide;  
KW gastric disorders; pancreatic disease; Schistosoma mansoni infection;  
KW spinocerebellar ataxia; Plasmodium falciparum parasitaemia; diabetes;  
KW Corneal dystrophy-Greponow type I; Corneal dystrophy-lattice type I;  
KW Reis-Bucklers corneal dystrophy; cytostatic; immunosuppressive;  
KW antiallergic; antiaesthetic; antiinfectivity; antiinflammatory;  
KW antidiabetic; protozoacide; hepatotropic; virucide; ophthalmological;  
KW gynaecological; antiinfertility; immunostimulant; auditory; haemostatic;  
KW gene therapy; FCTR3b; neurestin-like protein.  
XX  
OS Homo sapiens.  
XX  
PN WO20016747-A2.  
XX  
PD 13-SEP-2001.  
XX  
PE 05-MAR-2001; 2001WO-US07160.  
XX  
PF 03-MAR-2000; 2000US-186592P.  
PR 03-MAR-2000; 2000US-186718P.  
PR 06-MAR-2000; 2000US-187293P.  
PR 06-MAR-2000; 2000US-187294P.  
PR 17-MAR-2000; 2000US-190400P.  
PR 07-APR-2000; 2000US-196018P.  
PR 03-JAN-2001; 2001US-259548P.  
XX  
PA (CURA-) CURAGEN CORP.  
XX

PI Vernet CAM, Fernandes E, Shimkets RA, Herrmann JL, Majumder K;  
PI Macdougall J, Mishra V, Mezes PS, Rastelli L;  
DR WPI: 2001-596637/67.  
DR N-PSDB; AAS14085.  
XX  
XX  
PT Novel polypeptides designated as FCTR<sub>X</sub> polypeptides, useful in  
PT detection, prevention and treatment of a broad range of pathological  
PT states -  
XX  
XX  
PS Claim 1; Page 35-36; 215pp; English.  
XX  
XX The invention relates to human FCTR<sub>X</sub> polypeptides, FCTR1-FCTR7, and the  
CC nucleic acids encoding them. These sequences are useful for the treatment  
CC or prevention of numerous disorders including myelogenous leukaemia,  
CC carcinomas, melanomas, gliomas, astrocytomas, congenital neonatal  
CC alloimmune thrombocytopenia, neurological disorders, neurodegenerative  
CC disorders, nerve trauma, familial myelodysplastic syndrome,  
CC Charcot-Marie-Tooth neuropathy, demyelinating Gardner syndrome, familial  
CC myelodysplastic syndrome, mental health conditions, immunological  
CC disorders, allergy and infection, bronchial asthma, Avelino type  
CC eosinophilia, lung diseases, reproductive disorders, infertility, male  
CC and female reproductive disorders, deafness, glycoprotein Ia deficiency,  
CC desmoid disease, Turcot syndrome, liver cirrhosis, hepatitis C, gastric  
CC disorders, pancreatic diseases such as diabetes, Schistosoma mansoni  
CC infection, spinocerebellar ataxia, Plasmodium falciparum parasitaemia,  
CC Corneal dystrophy-Greenow type I, Corneal dystrophy-lattice type I and  
CC Reis-Bucklers corneal dystrophy. This sequence represents FCTR3b, a  
CC neurastin-like protein.  
XX  
SQ Sequence 2733 AA;  
  
Query Match 6.6%; Score 7; DB 22; Length 2733;  
Best Local Similarity 100.0%; Pred. No. 4.5e+02;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
Qy 57 MVNLSG 63  
|||  
|||  
DB 2010 MVNLSG 2016  
  
RESULT 18  
AAW21483  
ID AAW21483 standard; peptide; 6 AA.  
XX  
XX AAW21483;  
AC  
XX  
DT 30-JUL-1997 (first entry)  
XX  
XX  
DE Schistosoma elastase precursor derived signal oligopeptide #3.  
XX  
XX Hydrophilic; signal oligopeptide; hydrophilicity maxima; vaccine; SIV;  
KW competitive inhibitor; feedback regulator; synthesis; gastrin precursor;  
KW charge; polarity; farnesyl synthetase; plasminogen activator inhibitor 1;  
KW hydroxymethylglutaryl coenzyme A reductase; glucagon precursor; rhesus;  
KW gonadoliberin precursor; plasminogen activator inhibitor 2; prorenin;  
KW Alzheimer amyloid A4; corticotropin releasing factor binding protein;  
KW apolipoprotein E; herpes virus 1 glycoprotein B; HSV1; human; OMVVS;  
KW herpes virus 2 glycoprotein B; HSV2; collagenase; apolipoprotein A;  
KW treponema pallidum membrane protein; TMPA; islet amyloid polypeptide;  
KW fibroblast MMP1; schistosoma elastase precursor; schistosomin;  
KW hepatitis delta antigen; rev protein; HIV; VILV; angiotensinogen.  
XX  
XX Schistosoma sp.  
OS  
XX  
XX W09519568-A1.  
FN  
XX  
XX 20-JUL-1995.  
PD  
XX  
XX 12-JAN-1995; 95WO-US00575.  
PF  
XX  
XX 14-JAN-1994; 94US-0182248.  
PR  
XX

PA (RATH/) RATH M.  
XX  
XX  
PI Rath M;  
XX  
DR WPI: 1995-263953/34.  
XX  
XX  
PT Identifying signal oligopeptide(s) in protein sequence(s) - shown as  
PT regions of max. hydrophilicity, used in modulating communication  
PT between protein(s)  
XX  
XX  
PS Claim 5; Page 70; 88pp; English.  
XX  
XX The sequences given in AAW21201-560 represent hydrophilic signal oligo-  
CC peptides. These signal oligopeptides are localised on the surface  
CC of the protein and are represented by the hydrophilicity maxima of  
CC the protein. These peptides are enriched in charged amino acids  
CC arranged with neutral spacer amino acids. The specific signal  
CC character of these oligopeptides is determined by a characteristic  
CC combination of conformation and charge within the signal sequence.  
CC These oligopeptides may be used as vaccines in the treatment of  
CC human disease, as competitive inhibitors to prevent or reduce the  
CC metabolic action or interaction of a selected protein by blocking  
CC its specific signal sequences, or as therapeutic agents to function  
CC as feedback regulators to reduce synthesis rate of a selected protein.  
CC These peptides may be modified by omitting one or more amino acids at  
CC the N- and/or C-terminal, by substituting one or more amino acids  
CC without consideration of charge and polarity, by substituting one or  
CC more amino acids with amino acid residues with similar charge and/or  
CC polarity, by omitting one or more amino acids or a combination of these.  
XX  
SQ Sequence 6 AA;  
  
Query Match 5.7%; Score 6; DB 16; Length 6;  
Best Local Similarity 100.0%; Pred. No. 7.8e+05;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
Qy 34 ARQRRP 39  
|||  
|||  
DB 1 ARQRRP 6  
  
RESULT 19  
AAG95984  
ID AAG95984 standard; Peptide; 10 AA.  
XX  
XX AAG95984;  
AC  
XX  
DT 18-SEP-2001 (first entry)  
XX  
XX  
DE Human complementary peptide, SEQ ID NO: 2178.  
XX  
XX Human; complementary peptide; ligand; drug discovery; drug design.  
XX  
XX Homo sapiens.  
OS  
XX  
XX W0200142277-A2.  
FN  
XX  
XX 14-JUN-2001.  
PD  
XX  
XX 13-DEC-2000; 2000WO-GB04776.  
PF  
XX  
XX 13-DEC-1999; 99GB-0029464.  
PR  
XX  
XX (PROT-) PROTEOM LTD.  
PA  
XX  
XX Roberts GW, Heal JR;  
PI  
XX  
XX WPI: 2001-408419/43.  
DR  
XX  
XX  
PT A set of peptide ligands consisting of specific complementary peptides  
PT to proteins encoded by genes of the human genome, useful in an assay  
PT for screening and identifying of one or more novel peptides which are  
PT drug candidates or pro-drugs -

XX Example 4; Page 354; 646pp; English.  
PS  
CC The invention relates to a set of complementary peptide ligands  
CC generated from the human genome. The complementary peptides  
CC interact with their relevant target proteins encoded in the human  
CC genome. They can be used as reagents in drug discovery and as lead  
CC ligands to facilitate drug design and development. The present  
CC sequence is a complementary peptide provided in the specification.  
SQ Sequence 10 AA;  
Query Match 5.7%; Score 6; DB 22; Length 10;  
Best Local Similarity 100.0%; Pred. No. 28;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 33 SARQR 38  
DB 5 SARQR 10  
RESULT 20  
AA66990  
ID AA66990 standard; Peptide; 15 AA.  
XX  
AC AA66990;  
XX  
DT 15-MAR-2000 (first entry)  
XX  
DE B.lentus protease peptide A5.  
XX  
KM Subtilisin; allergy; human; bacterium; protease; epitope; detergent;  
XX cosmetic; textile; pet food industry; debriement treatment.  
OS Bacillus lentus.  
XX  
FN WO953078-A2.  
XX  
PD 21-OCT-1999.  
XX  
PF 15-APR-1999; 99WO-US08177.  
XX  
PR 15-APR-1998; 98US-0060854.  
XX  
PA (GEMV ) GENENCOR INT INC.  
XX  
PI Estell D;  
XX  
DR WPI; 2000-061971/05.  
XX  
PT Modified non-human protease having reduced allergenicity -  
XX  
PS Example 2; Fig 7; 38pp; English.  
XX  
CC Peptides AA66983-Y67070 represent peptides derived from the Bacillus  
CC lentus subtilisin protein. The peptides were used for epitope mapping of  
CC the bacterial subtilisin. The invention relates to a method of reducing  
CC the allergenicity of a non-human protein, especially a bacterial  
CC protease such as subtilisin, by identifying an epitope on the non-human  
CC protein and replacing it by an analogous region from the human protein.  
CC The method is useful for producing proteins, including proteases, that  
CC are less likely to cause allergic reactions. Thus the invention can be  
CC employed for example in detergents, cosmetics, textile treatment, and  
CC pet food industries. The human subtilisin can be used pharmaceutically  
CC for debriement treatments.  
SQ Sequence 15 AA;  
Query Match 5.7%; Score 6; DB 21; Length 15;  
Best Local Similarity 100.0%; Pred. No. 41;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 19 SGVKVA 24

DB 9 SGVKVA 14  
RESULT 21  
AA66991  
ID AA66991 standard; Peptide; 15 AA.  
XX  
AC AA66991;  
XX  
DT 15-MAR-2000 (first entry)  
XX  
DE B.lentus protease peptide A4.  
XX  
KM Subtilisin; allergy; human; bacterium; protease; epitope; detergent;  
XX cosmetic; textile; pet food industry; debriement treatment.  
OS Bacillus lentus.  
XX  
FN WO953078-A2.  
XX  
PD 21-OCT-1999.  
XX  
PF 15-APR-1999; 99WO-US08177.  
XX  
PR 15-APR-1998; 98US-0060854.  
XX  
PA (GEMV ) GENENCOR INT INC.  
XX  
PI Estell D;  
XX  
DR WPI; 2000-061971/05.  
XX  
PT Modified non-human protease having reduced allergenicity -  
XX  
PS Example 2; Fig 7; 38pp; English.  
XX  
CC Peptides AA66983-Y67070 represent peptides derived from the Bacillus  
CC lentus subtilisin protein. The peptides were used for epitope mapping of  
CC the bacterial subtilisin. The invention relates to a method of reducing  
CC the allergenicity of a non-human protein, especially a bacterial  
CC protease such as subtilisin, by identifying an epitope on the non-human  
CC protein and replacing it by an analogous region from the human protein.  
CC The method is useful for producing proteins, including proteases, that  
CC are less likely to cause allergic reactions. Thus the invention can be  
CC employed for example in detergents, cosmetics, textile treatment, and  
CC pet food industries. The human subtilisin can be used pharmaceutically  
CC for debriement treatments.  
SQ Sequence 15 AA;  
Query Match 5.7%; Score 6; DB 21; Length 15;  
Best Local Similarity 100.0%; Pred. No. 41;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 19 SGVKVA 24  
DB 6 SGVKVA 11  
RESULT 22  
AA66992  
ID AA66992 standard; Peptide; 15 AA.  
XX  
AC AA66992;  
XX  
DT 15-MAR-2000 (first entry)  
XX  
DE B.lentus protease peptide A3.  
XX  
KM Subtilisin; allergy; human; bacterium; protease; epitope; detergent;  
XX cosmetic; textile; pet food industry; debriement treatment.  
XX

XX Bacillus lentus.  
 XX  
 FN MO9953078-A2.  
 XX  
 PD 21-OCT-1999.  
 XX  
 PF 15-APR-1999; 99WO-US08177.  
 XX  
 PR 15-APR-1998; 98US-0060854.  
 XX  
 PA (GEMV ) GENENCOR INT INC.  
 XX  
 PI Esteel D;  
 XX  
 DR WPI ; 2000-061971/05.  
 XX  
 PT Modified non-human protease having reduced allergenicity -  
 XX  
 PS Example 2; Fig 7; 38pp; English.  
 XX  
 CC Peptides AAY66983-Y67070 represent peptides derived from the Bacillus  
 CC lentus subtilisin protein. The peptides were used for epitope mapping of  
 CC the bacterial subtilisin. The invention relates to a method of reducing  
 CC the allergenicity of a non-human protein, especially a bacterial  
 CC protease such as subtilisin, by identifying an epitope on the non-human  
 CC protein and replacing it by an analogous region from the human protein.  
 CC The method is useful for producing proteins, including proteases, that  
 CC are less likely to cause allergic reactions. Thus the invention can be  
 CC employed for example in detergents, cosmetics, textile treatment, and  
 CC pet food industries. The human subtilisin can be used pharmaceutically  
 CC for debridement treatments.  
 XX  
 Sequence 15 AA;

Query Match	5.7%	Score 6	DB 21	Length 15
Best Local Similarity	100.0%	Pred. No. 41		
Matches 6	Conservative 0	Mismatches 0	Indels 0	Gaps 0

QY	19	SGVKVA	24
Db	3	SGVKVA	8

RESULT 23  
 AAY54627  
 ID AAY54627 standard; Protein; 15 AA.

AC	AAV54627;
XX	
DT	04-FEB-2000 (first entry)

DE B. lentus protease fragment peptide A5.

KM Subtilisin, protease variant; precursor protease; cleaning composition detergent; liquid soap application; dish-care formulation; animal feed

KM contact lens cleaning solution; peptide hydrolysis; waste treatment;

KM cosmetic formulation; fusion-cleavage enzyme; protein production.

OS *Bacillus lentus*.

PN WO9953038-A2.

PD 21-OCT-1999.

PF 14-APR-1999; 99WO-US08253.

PR 15-APR-1998; 98US-0060872.

PA (GEMV ) GENENCOR INT INC.

PI Estell DA, Harding FA;

AA WPI; 2000-013100/01.  
DR

XX New mutant proteins having lower allergenic response in humans useful  
PT in cleaning compositions, animal feed and treating textiles -  
PT  
XX  
XX Example 2, Fig 6a; 43pp; English.

This sequence is a fragment of the *Bacillus lentus* protease. The invention relates to protease variants comprising a substitution at least 1 of the positions in a precursor protease corresponding to K170D, Y171Q and/or S173D of the *Bacillus amyloliquefaciens* subtilisin. The protease variant can be used in a cleaning composition (e.g. in detergents, in bar or liquid soap application, dish-care formulations, or contact lens cleaning solutions), in animal feed and for treating textiles (e.g. treating wool to prevent felting). The protease variant may also be used for peptide hydrolysis, waste treatment, cosmetic (e.g. skin care) formulations, or as fusion-cleavage enzymes in protein production. The proteins (including enzymes) with reduced antigenicity may be used with significantly less danger of sensitization for the individuals exposed. Peptides which contain epitopes responsible for initially sensitizing an individual may be identified by measuring the proliferation of T-cells due to T-cell epitope recognition.

Seq	Sequence	15 AA;
1	...	...
2	...	...
3	...	...
4	...	...
5	...	...
6	...	...
7	...	...
8	...	...
9	...	...
10	...	...
11	...	...
12	...	...
13	...	...
14	...	...
15	...	...
16	...	...
17	...	...
18	...	...
19	...	...
20	...	...
21	...	...
22	...	...
23	...	...
24	...	...
25	...	...
26	...	...
27	...	...
28	...	...
29	...	...
30	...	...
31	...	...
32	...	...
33	...	...
34	...	...
35	...	...
36	...	...
37	...	...
38	...	...
39	...	...
40	...	...
41	...	...
42	...	...
43	...	...
44	...	...
45	...	...
46	...	...
47	...	...
48	...	...
49	...	...
50	...	...
51	...	...
52	...	...
53	...	...
54	...	...
55	...	...
56	...	...
57	...	...
58	...	...
59	...	...
60	...	...
61	...	...
62	...	...
63	...	...
64	...	...
65	...	...
66	...	...
67	...	...
68	...	...
69	...	...
70	...	...
71	...	...
72	...	...
73	...	...
74	...	...
75	...	...
76	...	...
77	...	...
78	...	...
79	...	...
80	...	...
81	...	...
82	...	...
83	...	...
84	...	...
85	...	...
86	...	...
87	...	...
88	...	...
89	...	...
90	...	...
91	...	...
92	...	...
93	...	...
94	...	...
95	...	...
96	...	...
97	...	...
98	...	...
99	...	...
100	...	...

Query Match	5.7%	Score 6;	DB 21;	Length 15;
Similarity	100.0%	Pred. No. 41;		
Best Local				
Matches	6;	Conservative	0;	Mismatches 0;
				Gaps 0.

QY	19	SGVKVA	24
Db	9	SGVKVA	14

RESULT 24  
AAY54628  
ID AAY54628 standard; Protein; 15 AA

AC	AAV54628;
XX	
DT	04-FEB-2000 (first entry)

DE B. lentus protease fragment peptide A4.

KM Subtilisin, protease variant; precursor protease; cleaning composition.  
KM detergent; liquid soap application; dish-care formulation; animal feed  
KM contact lens cleaning solution; peptide hydrolysis; waste treatment;  
KM cosmetic formulation; fusion-cleavage enzyme; protein production.

OS *Bacillus lentus*.

AA W09953038-A2  
PN

PD 21-OCT-1999.

AA 14-APR-1999; 99WO-US08253.  
PF

AA 15-APR-1998; 98US-0060872.  
PR

PA (GEMV ) GENENCOR INT INC.

PI Estell DA, Harding FA;

AA WPI; 2000-013100/01.  
DR

AA New mutant proteins having lower allergenic response in humans useful  
PT

## XXI IN CLEANING COMPOSITIONS, and

Example 2: fig 2a, top, English.

CC This sequence is a fragment of the *Bacillus lentus* protease. The  
CC invention relates to protease variants comprising a substitution at least  
CC 1 of the positions in a precursor protease corresponding to R170D, Y171Q  
CC and/or S173D of the *Bacillus amyloliquefaciens* subtilisin. The protease

CC variant can be used in a cleaning composition (e.g. in detergents, in bar  
CC or liquid soap application, dish-care formulations, or contact lens  
CC cleaning solutions), in animal feed and for treating textiles  
CC (e.g. treating wool to prevent felting). The protease variant may also be  
CC used for peptide hydrolysis, waste treatment, cosmetic (e.g. skin care)  
CC formulations, or as fusion-cleavage enzymes in protein production. The  
CC proteins (including enzymes) with reduced antigenicity may be used with  
CC significantly less danger of sensitization for the individuals exposed.  
CC Peptides which contain epitopes responsible for initially sensitising an  
CC individual may be identified by measuring the proliferation of T-cells  
CC due to T-cell epitope recognition.

XX Sequence 15 AA;  
SQ

Query Match 5.7%; Score 6; DB 21; Length 15;  
Best Local Similarity 100.0%; Pred. No. 41;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 19 SGVKVA 24  
|||||  
Db 6 SGVKVA 11

RESULT 25  
AAVS4629  
ID AAVS4629 standard; Protein; 15 AA.  
XX  
AC AAVS4629;  
XX  
DT 04-FEB-2000 (first entry)  
XX  
DE B. lentus protease fragment peptide A3.  
XX  
KM Subtilisin; protease variant; precursor protease; cleaning composition;  
KM detergent; liquid soap application; dish-care formulation; animal feed;  
KM contact lens cleaning solution; peptide hydrolysis; waste treatment;  
KM cosmetic formulation; fusion-cleavage enzyme; protein production.  
XX  
OS Bacillus lentus.  
XX  
PN WO953038-A2.  
XX  
PD 21-OCT-1999.  
XX  
PF 14-APR-1999; 99WO-US08253.  
XX  
PR 15-APR-1998; 98US-0060872.  
XX  
PA (GEMV ) GENENCOR INT INC.  
XX  
PI Estell DA, Harding FA;  
XX  
DR WPI; 2000-013100/01.  
XX  
PT New mutant proteins having lower allergenic response in humans useful  
PT in cleaning compositions, animal feed and treating textiles -  
XX  
XX Example 2; Fig 6a; 43pp: English.

CC This sequence is a fragment of the Bacillus lentus protease. The  
CC invention relates to protease variants comprising a substitution at least  
CC 1 of the positions in a precursor protease corresponding to K170D, Y171Q  
CC and/or S173D of the Bacillus amyloliquefaciens subtilisin. The protease  
CC variant can be used in a cleaning composition (e.g. in detergents, in bar  
CC or liquid soap application, dish-care formulations, or contact lens  
CC cleaning solutions), in animal feed and for treating textiles  
CC (e.g. treating wool to prevent felting). The protease variant may also be  
CC used for peptide hydrolysis, waste treatment, cosmetic (e.g. skin care)  
CC formulations, or as fusion-cleavage enzymes in protein production. The  
CC proteins (including enzymes) with reduced antigenicity may be used with  
CC significantly less danger of sensitization for the individuals exposed.  
CC Peptides which contain epitopes responsible for initially sensitising an  
CC individual may be identified by measuring the proliferation of T-cells

CC due to T-cell epitope recognition.

XX Sequence 15 AA;  
SQ

Query Match 5.7%; Score 6; DB 21; Length 15;  
Best Local Similarity 100.0%; Pred. No. 41;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 19 SGVKVA 24  
|||||  
Db 3 SGVKVA 8

RESULT 26  
AAU38529  
ID AAU38529 standard; Peptide; 15 AA.  
XX  
AC AAU38529;  
XX  
DT 18-DEC-2001 (first entry)  
XX  
DE Bacillus lentus savinase, peptide #8.  
XX  
KM Immunogenic; immunostimulant; immunosuppressive; T-cell epitope;  
KM immune response; hormone; vaccine; cytokine; therapeutic; savinase;  
KM hyper-allergenic; hypo-allergenic; cleaning composition; human;  
KM textile treatment; contact lens cleaning solution; waste treatment;  
KM cosmetic formulation; subtilisin; BPN.  
XX  
OS Bacillus lentus.  
XX  
PN WO200159130-A2.  
XX  
PD 16-AUG-2001.  
XX  
PF 22-JAN-2001; 2001WO-US02204.  
XX  
PR 08-FEB-2000; 2000US-0500135.  
XX  
PA (GEMV ) GENENCOR INT INC.  
XX  
PI Estell DA, Harding FA;  
XX  
DR WPI; 2001-607196/69.  
XX  
PT A variant of a polypeptide of interest comprising an altered T-cell  
PT epitope to produce a different immunogenic response useful in  
PT therapeutics, vaccines, textile treatments and cosmetics -  
XX  
XX Example 2; Figure 6A; 54pp: English.

CC The invention relates to a variant of a polypeptide of interest  
CC comprising an altered T-cell epitope to produce a different immunogenic  
CC response. A method for altering the immunogenicity of a protein  
CC especially an enzyme, a hormone, a factor, a vaccine, or cytokine  
CC is described. The protein produces an immune response as desired and is  
CC useful in therapeutics, vaccines and in forming hyper and hypo-allergenic  
CC compounds, e.g. cleaning compositions, textile treatments, contact lens  
CC cleaning solutions, waste treatment products and cosmetic formulations.  
CC Unlike antibody assays, the method determines immune sensitivity to an  
CC epitope prior to any sensitisation taking place. AAU38518-AAU38742  
CC represent the amino acid sequences of Bacillus lentus savinase  
CC and human subtilisin (BPN) T-cell epitopes, and related  
CC amino acid sequences as described in the method of the invention.

XX Sequence 15 AA;  
SQ

Query Match 5.7%; Score 6; DB 22; Length 15;  
Best Local Similarity 100.0%; Pred. No. 41;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 19 SGVKVA 24  
|||||

Db 9 SGVKVA 14

RESULT 27  
AAU38530  
ID AAU38530 standard; Peptide; 15 AA.  
XX  
AC AAU38530;  
XX  
DT 18-DEC-2001 (first entry)  
XX  
DE Bacillus lentus savinase, peptide #9.  
XX  
KW immunogenic; immunostimulant; immunosuppressive; T-cell epitope;  
KW immune response; hormone; vaccine; cytokine; therapeutic; savinase;  
KW hyper-allergenic; hypo-allergenic; cleaning composition; human;  
KW textile treatment; contact lens cleaning solution; waste treatment;  
KW cosmetic formulation; subtilisin; BPN.  
XX  
OS Bacillus lentus.  
XX  
PN WO200159130-A2.  
XX  
PD 16-AUG-2001.  
XX  
PF 22-JAN-2001; 2001WO-US02204.  
XX  
PR 08-FEB-2000; 2000US-0500135.  
XX  
PA (GEMV ) GENENCOR INT INC.  
XX  
PI Estell DA, Harding FA;  
XX  
DR WPI; 2001-607196/69.  
XX  
PT A variant of a polypeptide of interest comprising an altered T-cell  
PT epitope to produce a different immunogenic response useful in  
PT therapeutics, vaccines, textile treatments and cosmetics -  
XX  
PS Example 2; Figure 6A; 54pp; English.  
XX  
CC The invention relates to a variant of a polypeptide of interest  
CC comprising an altered T-cell epitope to produce a different immunogenic  
CC response. A method for altering the immunogenicity of a protein  
CC especially an enzyme, a hormone, a factor, a vaccine, or cytokine  
CC is described. The protein produces an immune response as desired and is  
CC useful in therapeutics, vaccines and in forming hyper and hypo-allergenic  
CC compounds, e.g. cleaning compositions, textile treatments, contact lens  
CC cleaning solutions, waste treatment products and cosmetic formulations.  
CC Unlike antibody assays, the method determines immune sensitivity to an  
CC epitope prior to any sensitisation taking place. AAU38518-AAU38742  
CC represent the amino acid sequences of Bacillus lentus savinase  
CC and human subtilisin (BPN) T-cell epitopes, and related  
CC amino acid sequences as described in the method of the invention.  
XX  
SQ Sequence 15 AA;

Query Match 5.7%; Score 6; DB 22; Length 15;  
Best Local Similarity 100.0%; Pred. No. 41;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 19 SGVKVA 24  
Db 6 SGVKVA 11

RESULT 28  
AAU38531  
ID AAU38531 standard; Peptide; 15 AA.  
XX  
AC AAU38531;  
XX  
DT 18-DEC-2001 (first entry)

XX  
DE Bacillus lentus savinase, peptide #10.  
XX  
KW immunogenic; immunostimulant; immunosuppressive; T-cell epitope;  
KW immune response; hormone; vaccine; cytokine; therapeutic; savinase;  
KW hyper-allergenic; hypo-allergenic; cleaning composition; human;  
KW textile treatment; contact lens cleaning solution; waste treatment;  
KW cosmetic formulation; subtilisin; BPN.  
XX  
OS Bacillus lentus.  
XX  
PN WO200159130-A2.  
XX  
PD 16-AUG-2001.  
XX  
PF 22-JAN-2001; 2001WO-US02204.  
XX  
PR 08-FEB-2000; 2000US-0500135.  
XX  
PA (GEMV ) GENENCOR INT INC.  
XX  
PI Estell DA, Harding FA;  
XX  
DR WPI; 2001-607196/69.  
XX  
PT A variant of a polypeptide of interest comprising an altered T-cell  
PT epitope to produce a different immunogenic response useful in  
PT therapeutics, vaccines, textile treatments and cosmetics -  
XX  
PS Example 2; Figure 6A; 54pp; English.  
XX  
CC The invention relates to a variant of a polypeptide of interest  
CC comprising an altered T-cell epitope to produce a different immunogenic  
CC response. A method for altering the immunogenicity of a protein  
CC especially an enzyme, a hormone, a factor, a vaccine, or cytokine  
CC is described. The protein produces an immune response as desired and is  
CC useful in therapeutics, vaccines and in forming hyper and hypo-allergenic  
CC compounds, e.g. cleaning compositions, textile treatments, contact lens  
CC cleaning solutions, waste treatment products and cosmetic formulations.  
CC Unlike antibody assays, the method determines immune sensitivity to an  
CC epitope prior to any sensitisation taking place. AAU38518-AAU38742  
CC represent the amino acid sequences of Bacillus lentus savinase  
CC and human subtilisin (BPN) T-cell epitopes, and related  
CC amino acid sequences as described in the method of the invention.  
XX  
SQ Sequence 15 AA;

Query Match 5.7%; Score 6; DB 22; Length 15;  
Best Local Similarity 100.0%; Pred. No. 41;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 19 SGVKVA 24  
Db 3 SGVKVA 8

RESULT 29  
AAU29824  
ID AAU29824 standard; peptide; 21 AA.  
XX  
AC AAU29824;  
XX  
DT 15-NOV-1999 (first entry)  
XX  
DE Pneumolysin (263-81) FN(263-81).  
XX  
KW T-cell epitope; meningococcal; pneumococcal; bactericidal;  
KW multi-oligosaccharide glycoconjugate bacterial meningitis vaccine;  
KW pneumolysin; immunogenic; immune response; vaccination; meningitis;  
KW tumour-related antigen; diagnosis; detection; anti-cancer.  
XX  
OS Synthetic.



PN W09942130-A1.  
XX  
XX 26-AUG-1999.  
XX  
XX 23-FEB-1999; 99WO-CA00157.  
XX  
XX 23-FEB-1998; 98US-0027956.  
XX  
XX (CONN-) CONNAUGHT LAB LTD.  
XX  
XX Chong P, Klein MH, Lindberg A;  
XX WPI; 1999-540273/45.  
XX  
XX Multivalent immunogenic molecule comprising carrier with T cell  
PT epitope and many carbohydrate fragments with B cell epitopes;  
PT particularly for vaccination against meningitis and diagnosis  
XX  
XX Example 6; Page 53; 83pp; English.  
XX  
XX The present invention describes a multivalent immunogenic molecule (I)  
CC comprising: (i) carrier (Ia) having at least one functional T-cell  
CC epitope; and (ii) many different carbohydrate fragments (Ib), all linked  
CC to (Ia) and each having at least one functional B-cell epitope. (Ia)  
CC increases the immunogenicity of (Ib). AAY29819 to AAY29826 represent  
CC potential T-cell epitopes from meningococcal and pneumococcal proteins,  
CC used in the simplification of the present invention. (I) are used to  
CC generate an immune response, specifically for protective vaccination  
CC against meningitis (Streptococcus pneumoniae or Neisseria meningitidis),  
CC but also against tumour-related antigens and antigens from other  
CC bacteria, e.g. Escherichia coli, Salmonella typhi, Streptococcus mutans,  
CC Cryptococcus neoformans, Klebsiella, Staphylococcus aureus or  
CC Pseudomonas aeruginosa; to detect, by complex formation, (i) reactive  
CC antibodies and to raise (Ib)-specific antibodies, either for diagnostic  
CC detection of the corresponding antigen in usual immunoassays or, if  
CC directed against tumour antigens, for conjugation to anticancer agents.  
CC The combination of T- and B-cell epitopes in a single vaccine results  
CC in a strong and long-lasting humoral immunity.  
XX  
SQ Sequence 21 AA;  
Query Match 5.7%; Score 6; DB 20; Length 21;  
Best Local Similarity 100.0%; Pred. No. 55;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 20 GYKXAP 25  
DB 8 GYKXAP 13  
RESULT 30  
ABB37761  
ID ABB37761 standard; Peptide; 30 AA.  
XX  
XX ABB37761;  
XX  
XX 04-FEB-2002 (first entry)  
XX  
XX Peptide #5267 encoded by human foetal liver single exon probe.  
XX  
XX Human; foetal liver; gene expression; single exon nucleic acid probe.  
XX  
XX Homo sapiens.  
XX  
XX W0200157277-A2.  
XX  
XX 09-AUG-2001.  
XX  
XX 30-JAN-2001; 2001WO-US00669.  
XX  
XX 04-FEB-2000; 2000US-0180312.  
XX  
XX 26-MAY-2000; 2000US-0207456.  
XX  
XX 30-JUN-2000; 2000US-0608408.

PR 03-AUG-2000; 2000US-0632366.  
PR 21-SEP-2000; 2000US-0234687.  
PR 27-SEP-2000; 2000US-0236359.  
PR 04-OCT-2000; 2000GB-0024263.  
XX  
XX (MOLE-) MOLECULAR DYNAMICS INC.  
XX  
XX Penn SG, Hanzel DK, Chen W, Rank DR;  
XX WPI; 2001-483447/52.  
XX  
XX Human genome-derived single exon nucleic acid probes useful for  
PT analyzing gene expression in human fetal liver -  
XX  
XX Claim 27; SEQ ID NO 30396; 639pp + sequence listing; English.  
XX  
XX The invention relates to a single exon nucleic acid probe for  
CC measuring human gene expression in a sample derived from human foetal  
CC liver. The single exon nucleic acid probes may be used for predicting,  
CC measuring and displaying gene expression in samples derived from human  
CC fetal liver. The present sequence is a peptide encoded by a single exon  
CC nucleic acid probe of the invention.  
CC Note: The sequence data for this patent did not form part of the  
CC printed specification, but was obtained in electronic format directly  
CC from WIPO at ftp.wipo.int/pub/published\_pct\_sequences.  
XX  
SQ Sequence 30 AA;  
Query Match 5.7%; Score 6; DB 22; Length 30;  
Best Local Similarity 100.0%; Pred. No. 77;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 92 NDRDPS 97  
DB 6 NDRDPS 11  
RESULT 31  
ABG40652  
ID ABG40652 standard; Peptide; 30 AA.  
XX  
XX ABG40652;  
XX  
XX 19-AUG-2002 (first entry)  
XX  
XX Human peptide encoded by genome-derived single exon probe SEQ ID 30317.  
XX  
XX Human; single exon probe; asthma; lung cancer; COPD; IHD;  
XX chronic obstructive pulmonary disease; interstitial lung disease;  
XX familial idiopathic pulmonary fibrosis; neurofibromatosis;  
XX tuberous sclerosis; Gaucher's disease; Niemann-Pick disease;  
XX Hermansky-Rudlak syndrome; sarcoidosis; pulmonary haemosiderosis;  
XX pulmonary histiocytosis; lymphangioleiomyomatosis; Karsenger syndrome;  
XX pulmonary alveolar proteinosis; fibrocystic pulmonary dysplasia;  
XX primary ciliary dyskinesia; pulmonary hypertension;  
XX hyaline membrane disease.  
XX  
XX Homo sapiens.  
XX  
XX W0200186003-A2.  
XX  
XX 15-NOV-2001.  
XX  
XX 30-JAN-2001; 2001WO-US00665.  
XX  
XX 04-FEB-2000; 2000US-180312P.  
XX  
XX 26-MAY-2000; 2000US-207456P.  
XX  
XX 30-JUN-2000; 2000US-0608408.  
XX  
XX 03-AUG-2000; 2000US-0632366.  
XX  
XX 21-SEP-2000; 2000US-234687P.  
XX  
XX 27-SEP-2000; 2000US-236359P.  
XX  
XX 04-OCT-2000; 2000GB-0024263.

PA (MOLE-) MOLECULAR DYNAMICS INC.  
XX  
PI Penn SG, Hanzel DK, Chen W, Rank DR;  
XX  
DR WPI; 2002-114163/15.  
XX  
PT Spatially-addressable set of single exon nucleic acid probes, used to  
PT measure gene expression in human lung samples -  
XX  
PS Claim 27; SEQ ID No 30317; 634bp; English.  
XX  
CC The invention relates to a spatially-addressable set of single exon  
CC nucleic acid probes for measuring gene expression in a sample derived  
CC from human lung comprising single exon nucleic acid probes having one of  
CC 12614 nucleic acid sequences mentioned in the specification, or their  
CC complements or the 12387 open reading frames derived from the 12614  
CC probes. Also included are a microarray comprising the novel set of  
CC probes; the novel set of probes which hybridize at high stringency to a  
CC nucleic acid expressed in the human lung; measuring gene expression in a  
CC sample derived from human lung, comprising (a) contacting the array with  
CC a collection of detectably labeled nucleic acids derived from human lung  
CC mRNA, and (b) measuring the labeled nucleic acids bound to each probe of  
CC the array; identifying exons in a eukaryotic genome, comprising  
CC (a) algorithmically predicting at least one exon from genomic sequences  
CC of the eukaryote; and (b) detecting specific hybridisation of detectably  
CC labeled nucleic acids from eukaryote lung mRNA, to a single exon probe,  
CC having a fragment identical to the predicted exon, the probe is included  
CC in the above mentioned microarray; assigning exons to a single gene,  
CC comprising (a) identifying exons from genomic sequence by the method  
CC above and (b) measuring the expression of each of the exons in several  
CC tissues and/or cell types using hybridisation to a single exon  
CC microarrays having a probe with the exon, where a common pattern of  
CC expression of the exons in the tissues and/or cell types indicates that  
CC the exons should be assigned to a single gene; a peptide comprising one  
CC of 12011 sequences, mentioned in the specification, or encoded by the  
CC probes/open reading frames (ORF). The probes are used for gene  
CC expression analysis, and for identifying exons in a gene, particularly  
CC using human lung derived mRNA and for the study of lung diseases  
CC such as asthma, lung cancer, chronic obstructive pulmonary disease  
CC (COPD), interstitial lung disease (ILD), familial idiopathic pulmonary  
CC fibrosis, neurofibromatosis, tuberous sclerosis, Gaucher's disease,  
CC Niemann-Pick disease, Hermansky-Pudlak syndrome, sarcoidosis, pulmonary  
CC haemosiderosis, pulmonary histiocytosis, lymphangioleiomyomatosis,  
CC pulmonary alveolar proteinosis, Karagener syndrome, fibrocystic  
CC pulmonary dysplasia, primary ciliary dyskinesia, pulmonary hypertension  
CC and hyaline membrane disease. The present sequence is a peptide/protein  
CC encoded by a single exon probe of the invention.  
CC Note: The sequence data for this patent did not form part  
CC of the printed specification, but was obtained in electronic  
CC format directly from WIPO at  
CC ftp.wipo.int/pub/published\_pct\_sequences.  
XX  
SQ Sequence 30 AA;  
  
Query Match 5.7%; Score 6; DB 23; Length 30;  
Best Local Similarity 100.0%; Pred. No. 77;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
QY 92 NDRDPS 97  
DB 6 NDRDPS 11  
  
RESULT 32  
ID ABB22985 standard; Protein; 34 AA.  
XX  
AC ABB22985;  
XX  
DT 23-JAN-2002 (first entry)  
XX  
DE Protein #4984 encoded by probe for measuring heart cell gene expression.  
XX

KW Human; gene expression; heart; microarray; vascular system;  
KW cardiovascular disease; hypertension; cardiac arrhythmia;  
KW congenital heart disease.  
XX  
OS Homo sapiens.  
XX  
PN W0200157274-A2.  
XX  
PD 09-AUG-2001.  
XX  
PF 30-JAN-2001; 2001WO-US00666.  
XX  
PR 04-FEB-2000; 2000US-0180312.  
XX  
PR 26-MAY-2000; 2000US-0207456.  
XX  
PR 30-JUN-2000; 2000US-0608408.  
XX  
PR 03-AUG-2000; 2000US-0632366.  
XX  
PR 21-SEP-2000; 2000US-0234687.  
XX  
PR 27-SEP-2000; 2000US-0236359.  
XX  
PR 04-OCT-2000; 2000GB-0024263.  
XX  
PA (MOLE-) MOLECULAR DYNAMICS INC.  
XX  
PI Penn SG, Hanzel DK, Chen W, Rank DR;  
XX  
DR WPI; 2001-488899/53.  
XX  
PT Single exon nucleic acid probes for analyzing gene expression in human  
PT hearts -  
XX  
PS Claim 15; SEQ ID No 24755; 530bp; English.  
XX  
CC The present invention relates to single exon nucleic acid probes for  
CC measuring human gene expression in a sample derived from human heart (see  
CC ABB21535-ABA1305). The present sequence is a protein encoded by one such  
CC probe. The probes may be used for predicting, measuring and displaying  
CC gene expression in samples derived from the human heart via microarrays.  
CC By measuring gene expression, the probes are useful for predicting,  
CC diagnosing, grading, staging, monitoring and prognosing diseases of the  
CC human heart and vascular system e.g. cardiovascular disease,  
CC hypertension, cardiac arrhythmias and congenital heart disease.  
CC Note: The sequence data for this patent did not form part of the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pct\_sequences.  
XX  
SQ Sequence 34 AA;  
  
Query Match 5.7%; Score 6; DB 22; Length 34;  
Best Local Similarity 100.0%; Pred. No. 87;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
QY 61 OSGIRV 66  
DB 8 OSGIRV 13  
  
RESULT 33  
ID AAM90227  
XX  
AC AAM90227 standard; Protein; 34 AA.  
XX  
DT 07-NOV-2001 (first entry)  
XX  
DE Human immune/haematopoietic antigen SEQ ID NO.17820.  
XX  
KW Human; immune; haematopoietic; immune/haematopoietic antigen; cancer;  
KW cytostatic; gene therapy; vaccine; metastasis.  
XX  
OS Homo sapiens.  
XX  
PN W0200157182-A2.  
XX  
PD 09-AUG-2001.

XX 17-JAN-2001; 2001WO-US01354.  
PF  
XX 31-JAN-2000; 2000US-0179065.  
PR 04-FEB-2000; 2000US-0180628.  
PR 24-FEB-2000; 2000US-0184664.  
PR 02-MAR-2000; 2000US-0186350.  
PR 16-MAR-2000; 2000US-0189874.  
PR 17-MAR-2000; 2000US-0190076.  
PR 18-APR-2000; 2000US-0198123.  
PR 19-MAY-2000; 2000US-0205515.  
PR 07-JUN-2000; 2000US-0209467.  
PR 28-JUN-2000; 2000US-0214886.  
PR 30-JUN-2000; 2000US-0215135.  
PR 07-JUL-2000; 2000US-0216647.  
PR 07-JUL-2000; 2000US-0216880.  
PR 11-JUL-2000; 2000US-0217487.  
PR 11-JUL-2000; 2000US-0217496.  
PR 14-JUL-2000; 2000US-0218290.  
PR 26-JUL-2000; 2000US-0220963.  
PR 26-JUL-2000; 2000US-0220964.  
PR 14-AUG-2000; 2000US-0224518.  
PR 14-AUG-2000; 2000US-0224519.  
PR 14-AUG-2000; 2000US-0225213.  
PR 14-AUG-2000; 2000US-0225214.  
PR 14-AUG-2000; 2000US-0225266.  
PR 14-AUG-2000; 2000US-0225267.  
PR 14-AUG-2000; 2000US-0225268.  
PR 14-AUG-2000; 2000US-0225270.  
PR 14-AUG-2000; 2000US-0225447.  
PR 14-AUG-2000; 2000US-0225757.  
PR 14-AUG-2000; 2000US-0225758.  
PR 14-AUG-2000; 2000US-0225759.  
PR 18-AUG-2000; 2000US-0226279.  
PR 22-AUG-2000; 2000US-0226681.  
PR 22-AUG-2000; 2000US-0226868.  
PR 22-AUG-2000; 2000US-0227182.  
PR 23-AUG-2000; 2000US-0227009.  
PR 30-AUG-2000; 2000US-0228924.  
PR 01-SEP-2000; 2000US-0229287.  
PR 01-SEP-2000; 2000US-0229344.  
PR 01-SEP-2000; 2000US-0229344.  
PR 01-SEP-2000; 2000US-0229345.  
PR 05-SEP-2000; 2000US-0229509.  
PR 05-SEP-2000; 2000US-0229513.  
PR 06-SEP-2000; 2000US-0230437.  
PR 06-SEP-2000; 2000US-0230438.  
PR 08-SEP-2000; 2000US-0231242.  
PR 08-SEP-2000; 2000US-0231243.  
PR 08-SEP-2000; 2000US-0231244.  
PR 08-SEP-2000; 2000US-0231413.  
PR 08-SEP-2000; 2000US-0231414.  
PR 08-SEP-2000; 2000US-0232080.  
PR 08-SEP-2000; 2000US-0232081.  
PR 12-SEP-2000; 2000US-0231968.  
PR 14-SEP-2000; 2000US-0232397.  
PR 14-SEP-2000; 2000US-0232398.  
PR 14-SEP-2000; 2000US-0232399.  
PR 14-SEP-2000; 2000US-0232400.  
PR 14-SEP-2000; 2000US-0232401.  
PR 14-SEP-2000; 2000US-0233063.  
PR 14-SEP-2000; 2000US-0233064.  
PR 14-SEP-2000; 2000US-0233065.  
PR 21-SEP-2000; 2000US-0234223.  
PR 21-SEP-2000; 2000US-0234274.  
PR 25-SEP-2000; 2000US-0234997.  
PR 25-SEP-2000; 2000US-0234998.  
PR 26-SEP-2000; 2000US-0235484.  
PR 27-SEP-2000; 2000US-0235834.  
PR 27-SEP-2000; 2000US-0235836.  
PR 29-SEP-2000; 2000US-0236327.  
PR 29-SEP-2000; 2000US-0236367.  
PR 29-SEP-2000; 2000US-0236368.

PR 29-SEP-2000; 2000US-0236369.  
PR 29-SEP-2000; 2000US-0236370.  
PR 02-OCT-2000; 2000US-0236802.  
PR 02-OCT-2000; 2000US-0237037.  
PR 02-OCT-2000; 2000US-0237038.  
PR 02-OCT-2000; 2000US-0237039.  
PR 02-OCT-2000; 2000US-0237040.  
PR 13-OCT-2000; 2000US-0239935.  
PR 13-OCT-2000; 2000US-0239937.  
PR 20-OCT-2000; 2000US-0240960.  
PR 20-OCT-2000; 2000US-0241221.  
PR 20-OCT-2000; 2000US-0241785.  
PR 20-OCT-2000; 2000US-0241786.  
PR 20-OCT-2000; 2000US-0241787.  
PR 20-OCT-2000; 2000US-0241808.  
PR 20-OCT-2000; 2000US-0241809.  
PR 20-OCT-2000; 2000US-0241826.  
PR 01-NOV-2000; 2000US-024617.  
PR 08-NOV-2000; 2000US-0246474.  
PR 08-NOV-2000; 2000US-0246475.  
PR 08-NOV-2000; 2000US-0246476.  
PR 08-NOV-2000; 2000US-0246477.  
PR 08-NOV-2000; 2000US-0246478.  
PR 08-NOV-2000; 2000US-0246523.  
PR 08-NOV-2000; 2000US-0246524.  
PR 08-NOV-2000; 2000US-0246525.  
PR 08-NOV-2000; 2000US-0246526.  
PR 08-NOV-2000; 2000US-0246527.  
PR 08-NOV-2000; 2000US-0246528.  
PR 08-NOV-2000; 2000US-0246532.  
PR 08-NOV-2000; 2000US-0246609.  
PR 08-NOV-2000; 2000US-0246610.  
PR 08-NOV-2000; 2000US-0246611.  
PR 08-NOV-2000; 2000US-0246613.  
PR 17-NOV-2000; 2000US-0249207.  
PR 17-NOV-2000; 2000US-0249208.  
PR 17-NOV-2000; 2000US-0249209.  
PR 17-NOV-2000; 2000US-0249210.  
PR 17-NOV-2000; 2000US-0249211.  
PR 17-NOV-2000; 2000US-0249212.  
PR 17-NOV-2000; 2000US-0249213.  
PR 17-NOV-2000; 2000US-0249214.  
PR 17-NOV-2000; 2000US-0249215.  
PR 17-NOV-2000; 2000US-0249216.  
PR 17-NOV-2000; 2000US-0249217.  
PR 17-NOV-2000; 2000US-0249218.  
PR 17-NOV-2000; 2000US-0249244.  
PR 17-NOV-2000; 2000US-0249245.  
PR 17-NOV-2000; 2000US-0249264.  
PR 17-NOV-2000; 2000US-0249265.  
PR 17-NOV-2000; 2000US-0249297.  
PR 17-NOV-2000; 2000US-0249299.  
PR 17-NOV-2000; 2000US-0249300.  
PR 01-DEC-2000; 2000US-0250160.  
PR 01-DEC-2000; 2000US-0250391.  
PR 05-DEC-2000; 2000US-0251030.  
PR 05-DEC-2000; 2000US-0251988.  
PR 05-DEC-2000; 2000US-0256719.  
PR 06-DEC-2000; 2000US-0251479.  
PR 08-DEC-2000; 2000US-0251856.  
PR 08-DEC-2000; 2000US-0251868.  
PR 08-DEC-2000; 2000US-0251869.  
PR 08-DEC-2000; 2000US-0251989.  
PR 08-DEC-2000; 2000US-0251990.  
PR 11-DEC-2000; 2000US-0254097.  
PR 05-JAN-2001; 2001US-0259678.  
PA (HUMA-) HUMAN GENOME SCI INC.  
XX  
XX  
PI Rosen CA, Barash SC, Ruden SM;  
XX WPI; 2001-483426/52.  
DR N-PSDB; AAK63008.

XX Nucleic acids encoding human immune/hematopoietic antigen polypeptides,  
 PT useful for preventing, diagnosing and/or treating cancers and  
 PT metastasis -  
 XX  
 PS Claim 11; SEQ ID NO 17820; 3071pp + Sequence Listing; English.  
 CC  
 CC AAK54951 to AAK64702 encode the human immune/haematopoietic antigen (I)  
 CC amino acid sequences given in AAM82170 to AAM91921. (I) have cytosolic  
 CC activity, and can be used in gene therapy and vaccine production. (I)  
 CC proteins and polynucleotides may be used in the prevention, diagnosis and  
 CC treatment of diseases associated with inappropriate (I) expression. For  
 CC example, they may be used to treat disorders associated with decreased  
 CC expression by rectifying mutations or deletions in a patient's genome  
 CC that affect the activity of (I) by expressing inactive proteins or to  
 CC supplement the patient's own production of (I). Additionally, (I)  
 CC polynucleotides may be used to produce the secreted (I), by inserting  
 CC the nucleic acids into a host cell and culturing the cell to express the  
 CC protein. (I) proteins and polynucleotides may be used to prevent,  
 CC diagnose and treat immune/haematopoietic-related diseases, especially  
 CC cancers and cancer metastases of haematopoietic-derived cells. AAK64703  
 CC to AAK87694 represent human immune/haematopoietic antigen genomic  
 CC sequences from the present invention. AAK54942 to AAK54950 and AAM82169  
 CC represent sequences used in the exemplification of the present invention.  
 CC  
 SQ Sequence 34 AA;  
 Query Match 5.7%; Score 6; DB 22; Length 34;  
 Best Local Similarity 100.0%; Pred. No. 87;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 99 KNGGIL 104  
 DB 5 KNGGIL 10  
 RESULT 34  
 AAE04671  
 ID AAE04671 standard; peptide; 39 AA.  
 AC AAE04671;  
 DT 04-SEP-2001 (first entry)  
 DE Haem protein related Roussetus aegyptiacus protein fragment.  
 DE  
 KW Haem binding protein; Hemat-Hs; Hemat-Bs; gaseous ligand sensor;  
 KW oxygen storage; artificial photocynthesis; signalling function;  
 KW alpha-haemoglobin; myoglobin; therapy.  
 OS Roussetus aegyptiacus.  
 OS  
 FH Key Location/Qualifiers  
 FT Domain 19..39  
 FT /Label= M2\_Box  
 PD  
 PD 07-JUN-2001.  
 PD  
 PF 05-DEC-2000; 2000WO-US33048.  
 PF  
 PR 06-DEC-1999; 99US-0455978.  
 PR  
 PA (UYHA-) UNIV HAWAII.  
 PA  
 PI Alam M, Larsen R;  
 PI  
 DR WPI; 2001-374832/39.  
 DR  
 PT Novel isolated bacterial heme binding protein, Hemat-Hs or Hemat-Bs  
 PT which reversibly binds oxygen with low affinity, useful for controlled  
 PT storage of oxygen and for sensing gaseous ligands such as oxygen -

XX Disclosure; Page 16; 94pp; English.  
 PS  
 XX The present invention relates to isolated archael and bacterial haem  
 CC binding protein, Hemat-Hs or Hemat-Bs which reversibly binds oxygen  
 CC with low affinity. Haem binding protein is useful for controlled storage  
 CC of oxygen by allowing haem binding protein to bind and store oxygen, and  
 CC triggering the release of oxygen from haem binding protein by activating  
 CC the signalling domain. Haem binding protein is useful for sensing gaseous  
 CC ligands such as oxygen, NO, CO, or CN. Blood substitute comprising haem  
 CC binding protein is useful for treating a patient suffering from low  
 CC blood levels by administering and regulating the oxygen binding of the  
 CC haem-binding protein by modifying the signalling domain. Haem binding  
 CC protein is useful for haem-based catalysis, for artificial  
 CC photocynthesis and for identifying potential signalling functions of  
 CC mutated alpha-haemoglobin and myoglobin causing several diseases. The  
 CC present sequence is a haem protein related Roussetus aegyptiacus protein  
 CC fragment.  
 CC  
 SQ Sequence 39 AA;  
 Query Match 5.7%; Score 6; DB 22; Length 39;  
 Best Local Similarity 100.0%; Pred. No. 98;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 101 GGLLKK 106  
 DB 8 GGLLKK 13  
 RESULT 35  
 AAG03372  
 ID AAG03372 standard; Protein; 51 AA.  
 AC AAG03372;  
 DT 06-OCT-2000 (first entry)  
 DE Human secreted protein, SEQ ID NO: 7453.  
 DE  
 KW Human; 5' EST; expressed sequence tag; secreted protein; cDNA isolation;  
 KW gene therapy; chromosome mapping.  
 KW  
 OS Homo sapiens.  
 OS  
 PN EP1033401-A2.  
 PN  
 PD 06-SEP-2000.  
 PD  
 PF 21-FEB-2000; 2000EP-0200610.  
 PF  
 PR 26-FEB-1999; 99US-0122487.  
 PR  
 PA (GEST ) GENSET.  
 PA  
 PI Dumas Milne Edwards J, Duclert A, Giordano J;  
 PI  
 DR WPI; 2000-500381/45.  
 DR  
 DR N-PSDB; AAC03378.  
 DR  
 XX New nucleic acid that is a 5' expressed sequence tag (5' EST) for  
 PT obtaining cDNAs and genomic DNAs that correspond to 5'ESTs and for  
 PT diagnostic, forensic, gene therapy and chromosome mapping procedures -  
 XX  
 PS Claim 13; SEQ ID 7453; 71pp + CD-ROM; English.  
 CC  
 CC The present sequence is a polypeptide encoded by one of a large number  
 CC of 5' ESTs derived from mRNAs encoding secreted proteins. The 5' ESTs  
 CC were prepared from total human RNAs or polyA+ RNAs derived from 30  
 CC different tissues. EST sequences usually correspond mainly to the 3'  
 CC untranslated region (UTR) of the mRNA because they are often obtained  
 CC from oligo-dT primed cDNA libraries. Such ESTs are not well suited for  
 CC isolating cDNA sequences derived from the 5' ends of mRNAs and even in

CC those cases where longer cDNA sequences have been obtained, the full 5'  
CC UTR is rarely included. 5' ESTs are derived from mRNAs with intact 5'  
CC ends and can therefore be used to obtain full length cDNAs and genomic  
CC DNA. 5' ESTs are also used in diagnostic, forensic, gene therapy and  
CC chromosome mapping procedures. They are used to obtain upstream  
CC regulatory sequences and to design expression and secretion vectors.

XX Sequence 51 AA;

Query Match 5.7%; Score 6; DB 21; Length 51;

Best Local Similarity 100.0%; Pred. No. 1.3e+02; Mismatches 0; Indels 0; Gaps 0;

Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 35 RQRRI 40

Db 46 RQRRI 51

RESULT 36

AAU41874

ID AAU41874 standard; Protein; 54 AA.

XX AAU41874;

XX 13-FEB-2002 (first entry)

XX Propionibacterium acnes immunogenic protein #2770.

XX SARHO syndrome; synovitis; acne; pustulosis; hypertosis; osteomyelitis;  
XX uveitis; endophthalmitis; bone; joint; central nervous system; ELISA;  
XX inflammatory lesion; acne vulgaris; enzyme linked immunosorbent assay;  
XX dermatological; osteopathic; neuroprotectant.

XX Propionibacterium acnes.

XX WO200181581-A2.

XX 01-NOV-2001.

XX 20-APR-2001; 2001WO-US12865.

XX 21-APR-2000; 2000US-199047P.

XX 02-JUN-2000; 2000US-208841P.

XX 07-JUL-2000; 2000US-216747P.

XX (CORI-) CORIXA CORP.

XX Skeiky YAM, Persing DH, Mitcham JU, Wang SS, Bhatia A;

XX L'maisonneuve J, Zhang Y, Jen S, Carter D;

XX WPI: 2001-616774/71.

XX N-PSDB; AAS59515.

XX Example 1, SEQ ID No 3069, 1069pp; English.

XX Sequences AAU39105-AAU68017 represent Propionibacterium acnes immunogenic  
CC polypeptides. The proteins and their associated DNA sequences are used in  
CC the treatment, prevention and diagnosis of medical conditions caused by  
CC P. acnes. The disorders include SARHO syndrome (synovitis, acne,  
CC pustulosis, hypertosis and osteomyelitis), uveitis and endophthalmitis.  
CC P. acnes is also involved in infections of bone, joints and the central  
CC nervous system, however it is particularly involved in the inflammatory  
CC lesions associated with acne vulgaris. A method for detecting the  
CC presence or absence of P. acnes in a patient comprises contacting a  
CC sample with a binding agent that binds to the proteins of the invention  
CC and determining the amount of bound protein in the sample. The  
CC polypeptides may be used as antigens in the production of antibodies  
CC specific for P. acnes proteins. These antibodies can be used to  
CC downregulate expression and activity of P. acnes polypeptides and

CC therefore treat P. acnes infections. The antibodies may also be used as  
CC diagnostic agents for determining P. acnes presence, for example, by  
CC enzyme linked immunosorbent assay (ELISA).  
CC Note: The sequence data for this patent did not form part of the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pct\_sequences.

XX Sequence 54 AA;

Query Match 5.7%; Score 6; DB 22; Length 54;

Best Local Similarity 100.0%; Pred. No. 1.3e+02; Mismatches 0; Indels 0; Gaps 0;

Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 18 PSGVKV 23

Db 46 PSGVKV 51

RESULT 37

AAR95436

ID AAR95436 standard; Protein; 60 AA.

XX AAR95436;

XX 17-SEP-1996 (first entry)

XX Apomyoglobin BETA segment.

XX Multifunctional protein; targeted hetero-association;  
XX protein assembly; antibody engineering; apomyoglobin;  
XX myoglobin.

XX Synthetic.

XX WO9613583-A2.

XX 09-MAY-1996.

XX 20-OCT-1995; 95WO-EP04117.

XX 20-OCT-1994; 94EP-0116558.

XX (MORP-) MORPHOSYS GES PROTEINOPTIMIERUNG MBH.

XX Lupas A, Pack P;

XX WPI: 1996-239496/24.

XX N-PSDB; AAT15270.

XX Targeted hetero-association of recombinant proteins to  
XX multi-functional complexes - useful for therapeutic and diagnostic  
XX purposes

XX Example 4; Fig 17; 34pp; English.

XX Human apomyoglobin (aMG) can be used as an assembly device for  
CC the targeted hetero-association of recombinant proteins. The aMG  
CC is segmented into ALPHA (AAR95435), BETA (AAR95436) and GAMMA (AAR95437)  
CC segments encoded by Mro1-AscI-HindIII gene cassettes (AAT15269-71,  
CC respectively). Up to 6 functional domains (enzymes, antibody  
CC fragments, etc.) can be assembled by a 3-fold segmented structure,  
CC 3 at the N-termini and 3 at the C-termini of the segments. The  
CC presence of haem stabilises the native-like aMG fold and can be  
CC used as a switch to influence the association constant of the  
CC multi-functional complex.

XX Sequence 60 AA;

Query Match 5.7%; Score 6; DB 17; Length 60;

Best Local Similarity 100.0%; Pred. No. 1.5e+02; Mismatches 0; Indels 0; Gaps 0;

Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 101 GGIKK 106

Db 28 GILKCK 33

## RESULT 38

AAU64135 ID AAB23748 standard; protein; 61 AA.

AC AAB23748;

DT 10-JAN-2001 (first entry)

DE Ppu DnaB intein amino acid sequence.

KW Intein; Int-n; Int-c; DnaB; encryption; trait; split gene.

OS Porphyra purpurea.

PN W0200052146-A2.

PD 08-SEP-2000.

PP 03-MAR-2000; 2000WO-US05448.

PR 05-MAR-1999; 99US-0122943.

PR 02-JUL-1999; 99US-0142299.

PR 10-NOV-1999; 99US-0164617.

PR 10-NOV-1999; 99US-0164618.

PA (MAXY-) MAXYGEN INC.

PI Patten PA, Lasser M, Yamamoto T, Carr B, Ness JE, Bermudez ER;

DR WPI; 2000-549566/50.

PT Unencrypting traits using splice gene sequences -

PS Disclosure; Fig 6B; 77pp; English.

CC The present invention describes methods of unencrypting trait encrypted

CC gene sequences to provide unencrypted RNA and proteins. The present

CC sequence represents an intein amino acid sequence which is used in the

CC exemplification of the present invention.

SC Sequence 61 AA;

OY 2 SFLTLR 7

Db 25 SFLTLR 30

## RESULT 39

AAU64135 ID AAU64135 standard; Protein; 62 AA.

AC AAU64135;

DT 27-FEB-2002 (first entry)

DE Propionibacterium acnes immunogenic protein #25031.

KW SAPHO syndrome; synovitis; acne; pustulosis; hypertosis; osteomyelitis;

KW uveitis; endophthalmitis; bone; joint; central nervous system; ELISA;

OS Propionibacterium acnes.

PN W0200161581-A2.

XX 01-NOV-2001.

XX 20-APR-2001; 2001WO-US12865.

XX 21-APR-2000; 2000US-199047P.

XX 02-JUN-2000; 2000US-208841P.

XX 07-JUL-2000; 2000US-216747P.

PA (CORI-) CORIXA CORP.

PI Skeiky YAM, Persing DH, Mitcham JU, Wang SS, Bhatia A;

PI L'maisonneuve J, Zhang Y, Jen S, Carter D;

DR WPI; 2001-616774/71.

DR N-PSDB; AAS59639.

PT Propionibacterium acnes polypeptides and nucleic acids useful for

PT vaccinating against and diagnosing infections, especially useful for

PT treating acne vulgaris -

PS Example 1; SEQ ID No 25330; 1069pp; English.

XX Sequences AAU39105-AAU68017 represent Propionibacterium acnes immunogenic

XX polypeptides. The proteins and their associated DNA sequences are used in

XX the treatment, prevention and diagnosis of medical conditions caused by

XX P. acnes. The disorders include SAPHO syndrome (synovitis, acne,

XX pustulosis, hypertosis and osteomyelitis), uveitis and endophthalmitis.

XX P. acnes is also involved in infections of bone, joints and the central

XX nervous system, however it is particularly involved in the inflammatory

XX lesions associated with acne vulgaris. A method for detecting the

XX presence or absence of P. acnes in a patient comprises contacting a

XX sample with a binding agent that binds to the proteins of the invention

XX and determining the amount of bound protein in the sample. The

XX polypeptides may be used as antigens in the production of antibodies

XX specific for P. acnes proteins. These antibodies can be used to

XX downregulate expression and activity of P. acnes polypeptides and

XX therefore treat P. acnes infections. The antibodies may also be used as

XX diagnostic agents for determining P. acnes presence, for example, by

XX enzyme linked immunosorbent assay (ELISA).

CC Note: The sequence data for this patent did not form part of the printed

CC specification, but was obtained in electronic format directly from WIPO

CC at ftp.wipo.int/pub/published\_pct\_sequences.

SC Sequence 62 AA;

OY 32 MSAROR 37

Db 11 MSAROR 16

## RESULT 40

AAU32172 ID AAU32172 standard; Protein; 62 AA.

AC AAU32172;

DT 18-DEC-2001 (first entry)

DE Novel human secreted protein #2663.

KW Human; vaccination; gene therapy; nutritional supplement;

KW stem cell proliferation; haematopoiesis; nerve tissue regeneration;

OS Homo sapiens.

PN W0200179449-A2.

PD 25-OCT-2001.  
 XX  
 PF 16-APR-2001; 2001WO-US08656.  
 XX  
 PR 18-APR-2000; 2000US-0552929.  
 PR 26-JAN-2001; 2001US-0770160.  
 XX  
 PA (HYSE-) HYSEQ INC.  
 XX  
 PI Tang YT, Liu C, Drmanac RT;  
 XX  
 DR WPI; 2001-611725/70.  
 XX  
 PT Nucleic acids encoding a range of human polypeptides, useful in genetic  
 PT vaccination, testing and therapy -  
 XX  
 PS Claim 20; Page 564-565; 765pp; English.  
 XX  
 CC The invention relates to novel human secreted polypeptides. The  
 CC polypeptides and antibodies to the polypeptides are useful for  
 CC determining the presence of or predisposition to a disease associated  
 CC with altered levels of polypeptide. The polypeptides are also useful for  
 CC identifying agents (agonists and antagonists) that bind to them. Cells  
 CC expressing the proteins are useful for identifying a therapeutic agent  
 CC for use in treatment of a pathology related to aberrant expression or  
 CC physiological interactions of the polypeptide. Vectors comprising  
 CC the nucleic acids encoding the polypeptides and cells genetically  
 CC engineered to express them are also useful for producing the proteins.  
 CC The proteins are useful in genetic vaccination, testing and  
 CC therapy, and can be used as nutritional supplements. They may be used to  
 CC increase stem cell proliferation; to regulate haematopoiesis; and in  
 CC bone, cartilage, tendon and/or nerve tissue growth or regeneration;  
 CC immune suppression and/or stimulation; as anti-inflammatory agents; and  
 CC in treatment of leukaemias. AAU29510-AAU33304 represent the amino acid  
 CC sequences of novel human secreted proteins of the invention.  
 XX  
 SQ Sequence 62 AA;  
 Query Match 5.7%; Score 6; DB 22; Length 62;  
 Best Local Similarity 100.0%; Pred. No. 1.5e+02;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 17 QPSGVK 22  
 |||||  
 Db 27 QPSGVK 32

Search completed: April 9, 2003, 13:12:08  
 Job time : 96.623 secs

**THIS PAGE BLANK (USPTO)**



GenCore version 5.1.4 p5.4578  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using BW model

Run on: April 9, 2003, 12:50:42 ; Search time 93.8361 Seconds  
(without alignments)  
232.757 Million cell updates/sec

Title: US-10-020-441-2

Perfect score: 552  
Sequence: 1 VSEFLTLMGDDQGIHQPSG.....YGRDDNDPDRKNGKILKK 106

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

1: SPREMBL\_21:\*  
2: sp\_archaea:\*  
3: sp\_bacteria:\*  
4: sp\_fungi:\*  
5: sp\_human:\*  
6: sp\_invertebrate:\*  
7: sp\_mammal:\*  
8: sp\_mhc:\*  
9: sp\_organelle:\*  
10: sp\_plant:\*  
11: sp\_rodent:\*  
12: sp\_virus:\*  
13: sp\_vertebrate:\*  
14: sp\_unclassified:\*  
15: sp\_virus:\*  
16: sp\_bacteriap:\*  
17: sp\_archaeap:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	519	94.0	274	5	Q26552
2	442	80.1	257	5	Q26553
3	119	21.6	421	5	Q95R56
4	108.5	19.7	237	6	Q29464
5	102.5	18.6	273	6	Q9XSM1
6	99	17.9	233	4	Q96R27
7	99	17.9	275	4	Q96R26
8	98	17.8	115	4	Q96L36
9	98	17.8	130	4	Q8TD16
10	98	17.8	166	4	Q9UQ18
11	98	17.8	242	4	Q96R25
12	96.5	17.5	298	5	Q8T4N4
13	96	17.4	176	11	Q8VDE7
14	95.5	17.3	246	11	Q91VB1
15	95.5	17.3	270	4	Q96QL8
16	94.5	17.1	974	13	Q90WD8

17	93	16.8	246	11	Q9EP90	Q9EP90 mus musculu
18	90.5	16.4	257	6	Q19023	Q19023 macaca mula
19	88.5	16.0	281	5	Q76898	Q76898 drosophila
20	85.5	15.5	269	11	Q9D7T9	Q9D7T9 mus musculu
21	85.5	15.5	269	11	Q9C052	Q9C052 mus musculu
22	84.5	15.3	247	11	Q70500	Q70500 rattus norv
23	83.5	15.1	268	5	Q17030	Q17030 anopheles g
24	83	15.0	273	11	Q921N4	Q921N4 mus musculu
25	82.5	14.9	422	4	Q8WVC1	Q8WVC1 homo sapien
26	82.5	14.9	978	5	P91777	P91777 pacifastacu
27	82	14.9	365	5	Q9Y1K7	Q9Y1K7 anopheles g
28	81.5	14.8	286	4	Q13209	Q13209 homo sapien
29	81	14.7	331	11	Q8RIA6	Q8RIA6 mus musculu
30	80.5	14.6	266	6	Q46644	Q46644 macaca fasc
31	80.5	14.6	470	5	Q8T3A1	Q8T3A1 ciona intes
32	80	14.5	200	11	Q924U6	Q924U6 mus musculu
33	80	14.5	248	5	Q9XYX9	Q9XYX9 rhyzopertha
34	80	14.5	1524	13	Q91674	Q91674 xenopus lae
35	79.5	14.4	260	13	Q9W7P9	Q9W7P9 paratichthy
36	79	14.3	239	6	Q9GME0	Q9GME0 ornithorhyn
37	78.5	14.2	57	4	Q14237	Q14237 homo sapien
38	78.5	14.2	236	11	Q921H1	Q921H1 mus musculu
39	78	14.1	321	4	Q96R28	Q96R28 homo sapien
40	77.5	14.1	365	5	Q97366	Q97366 holotrichia
41	77.5	14.0	255	3	Q9Y7A9	Q9Y7A9 metathizium
42	77.5	14.0	269	6	Q95KW7	Q95KW7 bos taurus
43	77	13.9	128	11	Q9E013	Q9E013 mus musculu
44	76.5	13.9	388	5	Q966V2	Q966V2 halocynthia
45	76	13.8	716	11	Q91XG8	Q91XG8 mus musculu

## ALIGNMENTS

RESULT 1  
ID Q26552 PRELIMINARY; PRT; 274 AA.  
AC Q26552;  
DT 01-NOV-1996 (TREMBLrel. 01, Created)  
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)  
DT 01-MAR-2002 (TREMBLrel. 20, Last annotation update)  
DE Elastase.  
OS Schistosoma mansoni (Blood fluke).  
OC Eukaryota; Metazoa; Platyhelminthes; Trematoda; Digenea; Strigoidida;  
OC Schistosomatidae; Schistosomatidae; Schistosoma.  
OX NCBI\_TaxID=6183;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=PUERTO RICO.  
RX MEDLINE=96362066; PubMed=8720180;  
RA Pierrot C., Capron A., Khalife J.;  
RT "Cloning and characterization of two genes encoding Schistosoma  
mansoni elastase.";  
RL Mol. Biochem. Parasitol. 75:113-117(1995).  
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE  
TRYPSIN FAMILY.  
CC EMBL; U31768; AAC46967.1; -.  
DR HSSP; P20231; IAAO.  
DR MEROPS; S01.144; -.  
DR InterPro; IPR001314; Chymotrypsin.  
DR InterPro; IPR001254; Ser protease\_Try.  
DR Pfam; PF00089; trypsin\_1.  
DR PRINTS; PRO0722; CHYMOTRYPSIN.  
DR SMART; SM00020; TRYP\_SPE; 1.  
DR PROSITE; PS00240; TRYPsin\_DOM; 1.  
DR PROSITE; PS00134; TRYPsin\_HIS; UNKNOWN\_1.  
KW Hydrolyase; Serine protease.  
SQ SEQUENCE 274 AA; 29586 MW; 53F21914FDDDA450 CRC64;

Query Match 94.0%; Score 519; DB 5; Length 274;  
Best Local Similarity 95.3%; Pred. No. 2,4e-52;  
Matches 101; Conservative 1; Mismatches 4; Gaps 0;

Qy 1 VSFTLRNGDQGGIHHSQGVKAVAGYMPSCMSARQRRPIAGTSGFDIAIWMALQMVNL 60  
 |||||  
 Db 84 VSFTLRNGDQGGIHHSQGVKAVAGYMPSCMSARQRRPIAGTSGFDIAIWMALQMVNL 143  
 |||||

Qy 61 QSGIRVLSLPQPSDIPPEGTGVFIYVGRDNDPRPSKNGIILKK 106  
 |||||  
 Db 144 QSGIRVLSLPQPSDIPPEGTGVFIYVGRDNDPRPSKNGIILKK 189  
 |||||

## RESULT 2

Q26553 PRELIMINARY; PRT; 257 AA.

AC Q26553;  
 DT 01-NOV-1996 (TREMBlrel. 01, Created)  
 DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)  
 DT 01-MAR-2002 (TREMBlrel. 20, Last annotation update)  
 DE Elastase.  
 OS Schistosoma mansoni (Blood fluke).  
 OC Eukaryota; Metazoa; Platyhelminthes; Trematoda; Digenea; Strigoidida;  
 OC Schistosomatidae; Schistosomatidae; Schistosoma.  
 OX NCBI\_TaxID=6183;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=PUERTO RICO;  
 RX MEDLINE=96362066; PubMed=8720180;  
 RA Pierrot C., Capron A., Khalife J.;  
 RT "Cloning and characterization of two genes encoding Schistosoma  
 mansoni elastase."  
 RL Mol. Biochem. Parasitol. 75:113-117(1995).  
 DR EMBL; U31769; AAC64968.1; -.  
 DR HSSP; P00766; 1GCD.  
 DR InterPro; IPR001254; Ser\_protease\_Try.  
 DR Pfam; PF00089; trypsin\_1.  
 DR SMART; SM00020; Tryp\_Spc; 1.  
 DR PROSITE; PS50240; TRYPSIN\_DOM; 1.  
 DR PROSITE; PS00134; TRYPSIN\_HIS; UNKNOWN\_1.  
 DR Hydrolyase; Serine protease.  
 KW SEQUENCE 257 AA; 27971 MM; 00F8872799F9DD64 CRC64;

Query Match 80.1%; Score 442; DB 5; Length 257;

Best Local Similarity 84.3%; Pred. No. 2.1e-43;  
 Matches 86; Conservative 4; Mismatches 12; Indels 0; Gaps 0;

Qy 1 VSFTLRNGDQGGIHHSQGVKAVAGYMPSCMSARQRRPIAGTSGFDIAIWMALQMVNL 60  
 |||||  
 Db 80 VSFTLRNGDQGGIHHSQGVKAVAGYMPSCMSARQRRPIAGTSGFDIAIWMALQMVNL 139  
 |||||

Qy 61 QSGIRVLSLPQPSDIPPEGTGVFIYVGRDNDPRPSKNGI 102  
 |||||  
 Db 140 QSGIRVLSLPQPSDIPPEGTGVFIYVGRDNDPRPSKNGI 181  
 |||||

## RESULT 3

Q95RS6 PRELIMINARY; PRT; 421 AA.

AC Q95RS6;  
 DT 01-DEC-2001 (TREMBlrel. 19, Created)  
 DT 01-DEC-2001 (TREMBlrel. 19, Last sequence update)  
 DT 01-MAR-2002 (TREMBlrel. 20, Last annotation update)  
 DE Ld13269P.  
 OS Drosophila melanogaster (Fruit fly).  
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
 OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
 OC Ephydroidea; Drosophilidae; Drosophila.  
 OX NCBI\_TaxID=7227;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=BERKELEY;  
 RA Stapleton M., Brokstein P., Hong L., Aghayani A., Carlson J.,  
 RA Champagne M., Chavet C., Dorsett V., Farfan D., Frise E., George R.,  
 RA Gonzalez M., Guatin H., Li P., Liao G., Miranda C., Mungall C.J.,  
 RA Nuno J., Pacleb J., Paragas V., Park S., Pohnanavong S., Wan K.,  
 RA Yu C., Lewis S.E., Rubin G.M., Ceiniker S.;

RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.

DR EMBL; AY061167; AAL28715.1; -.  
 DR InterPro; IPR001128; Cytochrome\_P450.  
 DR InterPro; IPR001254; Ser\_protease\_Try.  
 DR Pfam; PF00089; trypsin\_1.  
 DR PROSITE; PS00086; CYTOCHROME\_P450; UNKNOWN\_1.  
 DR PROSITE; PS50240; TRYPSIN\_DOM; 1.  
 DR PROSITE; PS00134; TRYPSIN\_HIS; UNKNOWN\_1.  
 KW Hydrolyase; Serine protease.  
 SQ SEQUENCE 421 AA; 45646 MM; 68630BECDD959E3 CRC64;

Query Match 21.6%; Score 119; DB 5; Length 421;

Best Local Similarity 34.7%; Pred. No. 1.8e-05;  
 Matches 33; Conservative 13; Mismatches 27; Indels 22; Gaps 4;

Qy 5 TLNRGD-----QGGIHHSQGVKAVAGYMPSCMSARQRRPIAGTSGFDIAIWMAL 55  
 |||||  
 Db 218 TVRAGEWDQTKMKRLPYQERSVQVILHPDY-----NRSIA-----YDFALVILS 264  
 |||||

Qy 56 QMVNLQSGIRVLSLPQPSDIPPEGTGVFIYVGRDND 90  
 |||||  
 Db 265 QPVLDDHINVICLPQDDIPQGNCTGCTGMD 299  
 |||||

## RESULT 4

Q29464 PRELIMINARY; PRT; 237 AA.

AC Q29464;  
 DT 01-NOV-1996 (TREMBlrel. 01, Created)  
 DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)  
 DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)  
 DE Trypase (EC 3.4.21.59) (Fragment).  
 OS Bos taurus (Bovine).  
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;  
 OC Bovidae; Bovinae; Bos.  
 OX NCBI\_TaxID=9913;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=LIVER CAPSULE;  
 RX MEDLINE=96203914; PubMed=8620861.  
 RA Palladino M., Gambacorta A., Fiorucci L., Mignogna G., Barra D.,  
 RA Ascoli F.;  
 RT "cDNA cloning and primary structure of trypase from bovine mast cells  
 and evidence of the expression of bovine pancreatic trypsin inhibitor  
 mRNA in the same cells."  
 RL Eur. J. Biochem. 237:100-105(1996).

CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE  
 TRYPSIN FAMILY.  
 DR EMBL; X94982; CAA64438.1; -.  
 DR HSSP; P20231; IAAO.  
 DR MEROPS; S01.118; -.  
 DR InterPro; IPR001314; Chymotrypsin.  
 DR InterPro; IPR001254; Ser\_protease\_Try.  
 DR Pfam; PF00089; trypsin\_1.  
 DR PRINTS; PRO0722; CHYMOTRYPSIN.  
 DR SMART; SM00020; Tryp\_Spc; 1.  
 DR PROSITE; PS50240; TRYPSIN\_DOM; 1.  
 DR PROSITE; PS00134; TRYPSIN\_HIS; UNKNOWN\_1.  
 KW Hydrolyase; Serine protease.  
 FT NON TER 1  
 SQ SEQUENCE 237 AA; 26550 MM; EA537A75294EFBA CRC64;

Query Match 19.7%; Score 108.5; DB 6; Length 237;

Best Local Similarity 33.3%; Pred. No. 0.00015;  
 Matches 22; Conservative 16; Mismatches 19; Indels 9; Gaps 1;

Qy 29 PSCMSARQRRPIAGTSGFDIAIWMALQMVNLQSGIRVLSLPQPSDIPPEGTGVFIYVGR 88  
 |||||  
 Db 73 PNCYSYK-----NKADIALLELDLVNISMHVQVTLPEBSERTPPGTGCVITMG 123  
 |||||

Qy 89 RDDNDR 94  
 |||||

Db 124 NVDNGR 129

## RESULT 5

09XSM1 PRELIMINARY; PRT; 273 AA.

AC 09XSM1; PRELIMINARY; PRT; 273 AA.  
DT 01-NOV-1999 (TREMBlrel. 12, Created)  
DT 01-NOV-1999 (TREMBlrel. 12, Last sequence update)  
DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)  
DE Ovis aries (sheep).  
OS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;  
OC Bovidae; Caprinae; Ovis.  
OX NCBI\_TaxId=9940;

RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=BONE MARROW; PubMed=10848900;  
RA MEDLINE=20308142; PubMed=10848900;  
RA Pemberton A.D., McAliese S.M., Huntley J.F., Collie D.D.S.,  
RA Scudamore C.L., McEuen A.R., Walls A.F., Miller H.R.P.;  
RT "cdna sequence of two sheep mast cell tryptases and the differential  
RT expression of tryptase and sheep mast cell proteinase-1 in lung,  
RT dermis and gastrointestinal tract."  
RL Clin. Exp. Allergy 30:818-883(2000).  
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1, ALSO KNOWN AS THE  
CC TRYPSIN FAMILY

DR EMBL; Y18223; CAB41988.1; -.  
DR HSSP; P20231; LAAO.  
DR MEROPS; S01.118; -.  
DR InterPro; IPR001314; Chymotrypsin.  
DR InterPro; IPR001254; Ser protease\_Try.  
DR Pfam; PF00089; trypsin; 1.  
DR PRINTS; PRO0722; CHYMOTRYPSIN.  
DR SMART; SM00020; Tryp\_Spc; 1.  
DR PROSITE; PS50240; TRYPSIN\_DOM; 1.  
DR PROSITE; PS00134; TRYPSIN\_HIS; UNKNOWN\_1.  
DR PROSITE; PS00135; TRYPSIN\_SER; 1.  
KM Hydrolase; Serine protease.  
PT CHAIN 29 273  
SQ SEQUENCE 273 AA; 30156 MW; 88A19DC779053898 CRC64;

Query Match 18.6%; Score 102.5; DB 6; Length 273;  
Best Local Similarity 25.2%; Pred. No. 0.00092;  
Matches 31; Conservative 23; Mismatches 36; Indels 33; Gaps 4;

QY 4 LTLRNGDQGGIHQPPSGVAVPGVM-----PSCMSARQR-----PI 40  
DB 44 VSLRVDDQVM-RHOCGGSLIHPOWVLTAAHCTGPELQSPSPRRVQLRQHLYYQDRLPI 102  
QY 41 AQLT-----SGFDIAIVMLAQMNVNLSGIRIVISPPSDDIPPPGTGVPIVYGGDD 91  
DB 103 SRVIRPHYVWENGADIALQLLEBPVSI SCHVRVTLPPASETPPPSCQCVVGTGMDVND 162  
QY 92 NDR 94  
DB 163 NGR 165

## RESULT 6

096R27 PRELIMINARY; PRT; 233 AA.

AC 096R27; PRELIMINARY; PRT; 233 AA.  
DT 01-DEC-2001 (TREMBlrel. 19, Created)  
DT 01-DEC-2001 (TREMBlrel. 19, Last sequence update)  
DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)  
DE Mast cell tryptase beta III.  
GN TRYPTASEB.

OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
OX NCBI\_TaxId=9606;

RN [1]  
RP SEQUENCE FROM N.A.

RX MEDLINE=21096910; PubMed=11157797;  
RA Daniels R.J., Peden J.F., Lloyd C., Horsley S.W., Clark K.,  
RA Tufarelli C., Kearney L., Buckle V.J., Doggett N.A., Flint J.,  
RA Higgs D.R.;  
RT "Sequence, structure and pathology of the fully annotated terminal 2  
RT Mb of the short arm of human chromosome 16."  
RL Hum. Mol. Genet. 10:339-352(2001).  
DR EMBL; AB006466; AAK61270.1; -.

DR MEROPS; S01.027; -.  
DR InterPro; IPR001254; Ser protease\_Try.  
DR Pfam; PF00089; trypsin; 1.  
DR PROSITE; PS50240; TRYPSIN\_DOM; 1.  
DR PROSITE; PS00134; TRYPSIN\_HIS; UNKNOWN\_1.  
KM Hydrolase; Serine protease.

SQ SEQUENCE 233 AA; 25874 MW; 246A96C71B2CE6D0 CRC64;

Query Match 17.9%; Score 99; DB 4; Length 233;  
Best Local Similarity 39.6%; Pred. No. 0.0019;  
Matches 19; Conservative 11; Mismatches 18; Indels 0; Gaps 0;

QY 46 GFDIAIVMLAQMNVNLSGIRIVISPPSDDIPPPGTGVPIVYGGDDND 93  
DB 119 GADIALLEBPVNVSSHVHTVTLPPASETPPPGMPGCVVGTGMDVND 166

## RESULT 7

096R26 PRELIMINARY; PRT; 275 AA.

AC 096R26; PRELIMINARY; PRT; 275 AA.  
DT 01-DEC-2001 (TREMBlrel. 19, Created)  
DT 01-DEC-2001 (TREMBlrel. 19, Last sequence update)  
DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)  
DE Tryptase I.  
GN TRYPTASEC.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
OX NCBI\_TaxId=9606;

RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=21096910; PubMed=11157797;  
RA Daniels R.J., Peden J.F., Lloyd C., Horsley S.W., Clark K.,  
RA Tufarelli C., Kearney L., Buckle V.J., Doggett N.A., Flint J.,  
RA Higgs D.R.;  
RT "Sequence, structure and pathology of the fully annotated terminal 2  
RT Mb of the short arm of human chromosome 16."  
RL Hum. Mol. Genet. 10:339-352(2001).  
DR EMBL; AB006466; AAK61271.1; -.

DR MEROPS; S01.242; -.  
DR InterPro; IPR001254; Ser protease\_Try.  
DR Pfam; PF00089; trypsin; 1.  
DR PROSITE; PS50240; TRYPSIN\_DOM; 1.  
DR PROSITE; PS00134; TRYPSIN\_HIS; UNKNOWN\_1.  
DR PROSITE; PS00135; TRYPSIN\_SER; UNKNOWN\_1.  
KM Hydrolase; Serine protease.

SQ SEQUENCE 275 AA; 30601 MW; 2F4AAAA7DDBE483E CRC64;

Query Match 17.9%; Score 99; DB 4; Length 275;  
Best Local Similarity 39.6%; Pred. No. 0.0024;  
Matches 19; Conservative 11; Mismatches 18; Indels 0; Gaps 0;

QY 46 GFDIAIVMLAQMNVNLSGIRIVISPPSDDIPPPGTGVPIVYGGDDND 93  
DB 119 GADIALLEBPVNVSSHVHTVTLPPASETPPPGMPGCVVGTGMDVND 166

## RESULT 8

096L36 PRELIMINARY; PRT; 115 AA.

ID 096L36  
AC 096L36; PRELIMINARY; PRT; 115 AA.  
DT 01-DEC-2001 (TREMBlrel. 19, Created)

```
DT 01-DEC-2001 (TREMBlrel. 19, Last sequence update)
DT 01-MAR-2002 (TREMBlrel. 20, Last annotation update)
DE Delta I trypcase (Fragment).
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Hunt J.E., Wang H.W., Thomas P.S., McNeill P.;
RT "Cloning and characterization of novel human trypcase cDNAs.";
RL Submitted (SEP-2001) to the EMBL/Genbank/DBJ databases.
DR EMBL; AY055427; AAL17874.1; -.
DR InterPro; IPR001254; Ser_protease_Try.
DR Pfam; PF00089; trypsin; 1.
DR PROSITE; PS50240; TRYPSIN_DOM; 1.
DR PROSITE; PS00134; TRYPSIN_HIS; UNKNOWN_1.
KW Hydrolyase; Serine protease.
FT NON_TER 1
FT NON_TER 1
SQ SEQUENCE 115 AA; 12994 MW; E56C3C592D01BBE9 CRC64;

Query Match 17.8%; Score 98; DB 4; Length 115;
Best Local Similarity 40.7%; Pred. No. 0.0011;
Matches 22; Conservative 11; Mismatches 19; Indels 2; Gaps 1;

Qy 40 IAGTSGFDIAIVMAQWNLQSGIRVSLPQPSDIPPGTGVFIYGYGRDND 93
Db 54 IIQT--GADILALEBEPVNISHHITVTLPPASSTFPFGMPCWVTGMGDVNN 105

RESULT 9
O8TD16 PRELIMINARY; PRT; 130 AA.
AC O8TD16;
DT 01-JUN-2002 (TREMBlrel. 21, Created)
DT 01-JUN-2002 (TREMBlrel. 21, Last sequence update)
DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
DE Delta I trypcase (Fragment).
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Hunt J.E., Wang H.W., Thomas P.S., McNeill H.P.;
RT "Cloning and characterization of novel human trypcase cDNAs.";
RL Submitted (SEP-2001) to the EMBL/Genbank/DBJ databases.
DR EMBL; AF421357; AAL86695.1; -.
FT NON_TER 1
FT NON_TER 1
SQ SEQUENCE 130 AA; 14765 MW; E2E59C434AF3260F CRC64;

Query Match 17.8%; Score 98; DB 4; Length 130;
Best Local Similarity 28.9%; Pred. No. 0.0013;
Matches 24; Conservative 16; Mismatches 23; Indels 20; Gaps 2;

Qy 31 CMSAQQR-----PIAQT-----SGFDIAIVMAQWNLQSGIRVSLP 70
Db 43 CNEPVQLRQHLVYDOLLVPSRIIVHPQFYIIQTGADIALALEBPNVISHHITVTL 102

Qy 71 QPSDIPPGTGVFIYGYGRDND 93
Db 103 PASSTFPFGMPCWVTGMGDVNN 125

RESULT 10
O9UD18 PRELIMINARY; PRT; 166 AA.
AC O9UD18;
DT 01-MAY-2000 (TREMBlrel. 13, Created)
DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)
DT 01-MAR-2002 (TREMBlrel. 20, Last annotation update)
```

```
DE Mast cell mMCP-7-like I protein (Fragment).
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA MEDLINE=99121069; PubMed=9920877;
RA Pallao M., Fejzo M.S., Shavesteh L., Blount J.L., Caughey G.H.;
RT "Characterization of genes encoding known and novel human mast cell
RT trypases on chromosome 16p13.3.";
RL J. Biol. Chem. 274:3355-3362(1999).
CC -1-SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE
CC TRYPSIN FAMILY.
DR EMBL; AF099147; AAD17661.1; -.
DR HSSP; P20231; IAAO.
DR MEROPS; S01.054; -.
DR InterPro; IPR001314; Chymotrypsin.
DR InterPro; IPR001254; Ser_protease_Try.
DR Pfam; PF00089; trypsin; 1.
DR PRINTS; PR00722; CHYMOTRYPSIN.
DR SMART; SM00020; TRYD_SPC; 1.
DR PROSITE; PS50240; TRYPSIN_DOM; 1.
DR PROSITE; PS00134; TRYPSIN_HIS; UNKNOWN_1.
KW Hydrolyase; Serine protease.
FT NON_TER 166
FT NON_TER 166
SQ SEQUENCE 166 AA; 18433 MW; A7AF897A6423E7D5 CRC64;

Query Match 17.8%; Score 98; DB 4; Length 166;
Best Local Similarity 40.7%; Pred. No. 0.0017;
Matches 22; Conservative 11; Mismatches 19; Indels 2; Gaps 1;

Qy 40 IAGTSGFDIAIVMAQWNLQSGIRVSLPQPSDIPPGTGVFIYGYGRDND 93
Db 115 IIQT--GADILALEBEPVNISHHITVTLPPASSTFPFGMPCWVTGMGDVNN 166

RESULT 11
O96R25 PRELIMINARY; PRT; 242 AA.
AC O96R25;
DT 01-DEC-2001 (TREMBlrel. 19, Created)
DT 01-DEC-2001 (TREMBlrel. 19, Last sequence update)
DT 01-MAR-2002 (TREMBlrel. 20, Last annotation update)
DE Putative mast cell mMCP-7-like II trypcase (mMCP-7-like-2).
DN mMCP-7L.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA MEDLINE=21096910; PubMed=1157797;
RA Daniels R.J., Peden J.F., Lloyd C., Horsley S.W., Clark K.,
RA Tufarelli C., Kearney L., Buckle V.J., Doggett N.A., Flint J.,
RA Higgs D.R.;
RT "Sequence, structure and pathology of the fully annotated terminal 2
RT MD of the short arm of human chromosome 16.";
RL Hum. Mol. Genet. 10:339-352(2001).
DR EMBL; AE006466; AAK61272.1; -.
DR InterPro; IPR001254; Ser_protease_Try.
DR Pfam; PF00089; trypsin; 1.
DR PROSITE; PS50240; TRYPSIN_DOM; 1.
DR PROSITE; PS00134; TRYPSIN_HIS; UNKNOWN_1.
DR PROSITE; PS00135; TRYPSIN_SER; UNKNOWN_1.
KW Hydrolyase; Serine protease.
SQ SEQUENCE 242 AA; 26583 MW; 6E72D8AE1EDA2F1 CRC64;

Query Match 17.8%; Score 98; DB 4; Length 242;
Best Local Similarity 40.7%; Pred. No. 0.0027;
Matches 22; Conservative 11; Mismatches 19; Indels 2; Gaps 1;

Qy 40 IAGTSGFDIAIVMAQWNLQSGIRVSLPQPSDIPPGTGVFIYGYGRDND 93
```

Db 122 IIQT--GADIALBLEEPNISHTVTLPPASETFPPGMPGCVTGWGDVNN 173

## RESULT 12

08T4N4 PRELIMINARY; PRT; 298 AA.

AC 08T4N4;  
DT 01-JUN-2002 (Tremblrel. 21, Created)  
DT 01-JUN-2002 (Tremblrel. 21, Last sequence update)  
DT 01-JUN-2002 (Tremblrel. 21, Last annotation update)  
DE Midgut serine proteinase-1.  
OS Rhipicephalus appendiculatus (Brown ear tick).  
OC Eukaryota; Metazoa; Arthropoda; Chelicerata; Arachnida; Acari;  
OC Parasitiformes; Ixodida; Ixodidae; Rhipicephalus.  
OX NCBI\_TaxID=34631;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Mlenga A., Onuma M., Sugimoto C.;  
RT "Rhipicephalus appendiculatus midgut serine proteinase-1, cDNA cloning and characterization."  
RL Submitted (FEB-2002) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AY078093; AAL79565.1; -  
SQ SEQUENCE 298 AA; 32281 MW; 275E7064F0A5090D CRC64;

Query March 17.5%; Score 96.5; DB 5; Length 298;  
Best Local Similarity 30.0%; Pred. No. 0.0051;  
Matches 27; Conservative 11; Mismatches 39; Indels 13; Gaps 3;

QY 16 HOPSGVAVAFSGYMS--CMSARQRPPIAQTLSGFDIAIWMQVNLQSGIRVISLPP 72

Db 106 HSPCSRDTVEQYGAHLCMHVDPER-----DIGIVLKSSVNFDTVRPACLPBP 156

QY 73 SDIPEPGVFIVGGRDNDPRPKNG 102  
Db 157 GSELADGTLTYTGWGQTD-ETDPSSRADG 185

## RESULT 13

08VDE7 PRELIMINARY; PRT; 176 AA.

AC 08VDE7;  
DT 01-MAR-2002 (Tremblrel. 20, Created)  
DT 01-MAR-2002 (Tremblrel. 20, Last sequence update)  
DT 01-JUN-2002 (Tremblrel. 21, Last annotation update)  
DE Chymase 2.  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.  
OX NCBI\_TaxID=10090;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Chu W., Johnson D.A., Musich P.R.;  
RT "Molecular cloning and characterization of mouse mast cell chymases."  
RL Submitted (JAN-1992) to the EMBL/GenBank/DBJ databases.  
RN [2]  
RP SEQUENCE FROM N.A.  
RA Chu W.;  
RT "Molecular cloning and characterization of mouse mast cell chymases."  
RL Thesis (1991), Unknown Institution.

RN [3]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=92287966; PubMed=1376147;  
RA Chu W., Johnson D.A., Musich P.R.;  
RT "Molecular cloning and characterization of mouse mast cell chymases."  
RL Biochim. Biophys. Acta 1121:83-87(1992).  
DR EMBL; M68899; AAA39493.1; -  
DR InterPro; IPR001314; Chymotrypsin.  
DR InterPro; IPR001254; Ser.protease\_Try.  
DR Pfam; PF00089; trypsin\_1.  
DR PRINTS; PR00722; CHYMOTRYPSIN.  
DR SMART; SM00020; Tryp\_Spc; 1.  
DR PROSITE; PS50240; TRYPSIN\_DOM; 1.

DR PROSITE; PS00135; TRYPSIN\_SER; UNKNOWN\_1.  
SQ SEQUENCE 176 AA; 19173 MW; DCF586204DD18475 CRC64;

Query Match 17.4%; Score 96; DB 11; Length 176;  
Best Local Similarity 39.2%; Pred. No. 0.0031;  
Matches 20; Conservative 8; Mismatches 23; Indels 0; Gaps 0;

QY 48 DIAIWMQVNLQSGIRVISLPPGSDIPPGTGVFIVGGRDNDPRPSR 98  
Db 39 DIVLKLKKQNLTSADVVPPLPAPSDPAKGTGTCMAAGMGRGTGLKKSISR 89

## RESULT 14

091VB1 PRELIMINARY; PRT; 246 AA.

AC 091VB1;  
DT 01-DEC-2001 (Tremblrel. 19, Created)  
DT 01-DEC-2001 (Tremblrel. 19, Last sequence update)  
DT 01-JUN-2002 (Tremblrel. 21, Last annotation update)  
DE Mast cell protease 10.  
GN MCP10.  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
OX NCBI\_TaxID=10090;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STPAIN=BAIB/C, AND BAIB/CJ; TISSUE=JEJUNUM;  
RA Hunt J.E., McNeil P., Austen F., Gurish M., Stevens R.L.;  
RT "Molecular cloning of the cDNA and gene encoding mouse mast cell protease 10."  
RL Submitted (MAR-2001) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AF361940; AAK51075.1; -  
DR EMBL; AF361939; AAK51074.1; -  
DR MEROPS; S01.304; -  
DR InterPro; IPR001254; Ser.protease\_Try.  
DR Pfam; PF00089; trypsin\_1.  
DR PROSITE; PS50240; TRYPSIN\_DOM; 1.  
DR PROSITE; PS00134; TRYPSIN\_HIS; UNKNOWN\_1.  
DR PROSITE; PS00135; TRYPSIN\_SER; UNKNOWN\_1.  
KM Hydrolyase; Protease; Serine protease.  
SQ SEQUENCE 246 AA; 26857 MW; 3E2005E0E809FDD3 CRC64;

Query March 17.4%; Score 96; DB 11; Length 246;  
Best Local Similarity 39.2%; Pred. No. 0.0047;  
Matches 20; Conservative 8; Mismatches 23; Indels 0; Gaps 0;

QY 48 DIAIWMQVNLQSGIRVISLPPGSDIPPGTGVFIVGGRDNDPRPSR 98  
Db 109 DIVLKLKKQNLTSADVVPPLPAPSDPAKGTGTCMAAGMGRGTGLKKSISR 159

## RESULT 15

096QL8 PRELIMINARY; PRT; 270 AA.

AC 096QL8;  
DT 01-DEC-2001 (Tremblrel. 19, Created)  
DT 01-DEC-2001 (Tremblrel. 19, Last sequence update)  
DT 01-JUN-2002 (Tremblrel. 21, Last annotation update)  
DE Similar to elastase 3, pancreatic (protease E).  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=PANCREAS;  
RA Strauberg R.;  
RL Submitted (MAY-2001) to the EMBL/GenBank/DBJ databases.  
DR EMBL; BC008383; AAH08383.1; -  
DR MEROPS; S01.154; -  
DR InterPro; IPR001254; Ser.protease\_Try.  
DR Pfam; PF00089; trypsin\_1.



FT NON TER 257 257  
SQ SEQUENCE 257 AA; 27687 MW; 4D443DB67233D8DC CRC64;  
Query March 16.4%; Score 90.5; DB 6; Length 257;  
Best Local Similarity 27.9%; Pred. No. 0.021; Mismatches 35; Indels 11; Gaps 1;  
Matches 24; Conservative 16; Mismatches 35; Indels 11; Gaps 1;  
QY 4 LTRNGDQGIHOPSGVAVAGYMPSCMSARQRRPIAQTLSGFDAIYVLAQVNLQSG 63  
DB 77 LAVKGPQVPIPIINGDILFVHPLMRLCYAC-----GNDIALIKLSRSAGLQDA 125  
QY 64 IRVLSLPQSDIPPPGTGVFIYGYGR 89  
DB 126 VQLASLPAGDILPNETPCYITGMGR 151  
RESULT 19  
ID 076898 PRELIMINARY; PRT; 281 AA.  
AC 076898;  
DT 01-NOV-1998 (TREMBlrel. 08, Created)  
DT 01-NOV-1999 (TREMBlrel. 12, Last sequence update)  
DT 01-MAR-2002 (TREMBlrel. 20, Last annotation update)  
DE EG:80H7.1.  
GN Drosophila melanogaster (Fruit fly).  
OS Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
OC Eukaryota; Metazoa; Arthropoda; Insecta; Diptera; Brachycera; Muscomorpha;  
OC Ephydroidea; Drosophilidae; Drosophila.  
OX NCBI\_TaxID=7227;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Mottler S., Cadieu B., Dreano S., Lelaure V., Galibert F.;  
RT "Sequencing the distal X chromosome of Drosophila melanogaster.";  
RL Submitted (JUL-1998) to the EMBL/GenBank/DBJ databases.  
RN [2]  
RP SEQUENCE FROM N.A.  
RA Benos P.;  
RL Submitted (APR-1999) to the EMBL/GenBank/DBJ databases.  
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE  
TRYPSIN FAMILY.  
DR EMBL; AL031027; CAAL9843.2; -.  
DR HSSP; P00766; ICHG.  
DR FLYBASE; FBgn0025385; EG:80H7.1.  
DR InterPro; IPR001254; Chymotrypsin.  
DR InterPro; IPR001254; Ser\_protease\_Try.  
DR Pfam; PF00089; Trypsin\_1.  
DR PRINTS; PR00722; CHYMOTRYPSIN.  
DR SMART; SM00020; TRYD\_SPC; 1.  
DR PROSITE; PS50240; TRYPSIN\_DOM; 1.  
DR PROSITE; PS00134; TRYPSIN\_HIS; UNKNOWN\_1.  
DR PROSITE; PS00135; TRYPSIN\_SER; 1.  
KW Hydrolyase; Serine protease.  
SQ SEQUENCE 281 AA; 29875 MW; 4E78B0A60391949B CRC64;  
Query March 16.0%; Score 88.5; DB 5; Length 281;  
Best Local Similarity 24.1%; Pred. No. 0.041; Mismatches 40; Indels 25; Gaps 3;  
Matches 26; Conservative 17; Mismatches 40; Indels 25; Gaps 3;  
QY 4 LTRNGDQGIHOPSGVAVAGYMPSCMSARQRRPIAQTLSGFDAIYVLAQVNLQSG 58  
DB 79 LDLYGSGQMLARNSQVAVVAIFVHPCYEPEDKTVN-----DIALIQLAQSV 126  
QY 59 NLQSGIRIVISLPQSDIPPPGTGVFIYGYGRDNDNRDPSRKNGGLIKK 106  
DB 127 ALSKRVQGVRLPEPRQVTPGNASAVLAGWGLN-----ATGQVVOQ 166  
RESULT 20  
ID 09D779 PRELIMINARY; PRT; 269 AA.  
AC 09D779;  
DT 01-JUN-2001 (TREMBlrel. 17, Created)

DT 01-JUN-2001 (TREMBlrel. 17, Last sequence update)  
DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)  
DE 2310074F01RIK protein.  
GN ELA3B OR 2310074F01RIK.  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
OX NCBI\_TaxID=10090;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=C57BL/6J; TISSUE=STOMACH;  
RX MEDLINE=21085660; Pubmed=11217851;  
RA Kawai Y., Shingawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,  
RA Arakawa T., Hara A., Fukunishi Y., Kono H., Adachi J., Fukuda S.,  
RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanka I.,  
RA Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,  
RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,  
RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochwa H.,  
RA Kuehl P., Lewis S., Matsuo Y., Nakado I., Pesole G., Quackenbush J.,  
RA Schirini L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,  
RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,  
RA Blake J., Boffelli D., Bojunga N., Carinci P., de Bonaldo M.F.,  
RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,  
RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,  
RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,  
RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,  
RA Saeki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,  
RA Suzuki H., Toyooka K., Wang K.H., Weitz C., Whitaker C., Wilming L.,  
RA Wyshaw-Boriss A., Yoshida K., Hasegawa Y., Kawai H., Kontaiki S.,  
RA Hayashizaki Y.;  
RL "Functional annotation of a full-length mouse cDNA collection.";  
RT Nature 409:685-690(2001).  
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE  
TRYPSIN FAMILY.  
DR EMBL; AK008858; BAB25932.1; -.  
DR HSSP; P05805; IFON.  
DR MEROPS; S01.154; -.  
DR MGD; MGI:1915118; ELA3B.  
DR InterPro; IPR001314; Chymotrypsin.  
DR InterPro; IPR001254; Ser\_protease\_Try.  
DR Pfam; PF00089; Trypsin\_1.  
DR PRINTS; PR00722; CHYMOTRYPSIN.  
DR SMART; SM00020; TRYD\_SPC; 1.  
DR PROSITE; PS50240; TRYPSIN\_DOM; 1.  
DR PROSITE; PS00134; TRYPSIN\_HIS; UNKNOWN\_1.  
DR PROSITE; PS00135; TRYPSIN\_SER; 1.  
KW Hydrolyase; Serine protease.  
SQ SEQUENCE 269 AA; 28977 MW; 9F43F769DB2A7CF CRC64;  
Query March 15.5%; Score 85.5; DB 11; Length 269;  
Best Local Similarity 27.4%; Pred. No. 0.087; Mismatches 36; Indels 11; Gaps 1;  
Matches 23; Conservative 14; Mismatches 36; Indels 11; Gaps 1;  
QY 6 LTRNGDQGIHOPSGVAVAGYMPSCMSARQRRPIAQTLSGFDAIYVLAQVNLQSGIR 65  
DB 91 VEEGGEQVPIPIINGDILFVHPCYEPEDKTVN-----GNDIALIKLSRSAGLQDAVQ 139  
QY 66 VISLPQSDIPPPGTGVFIYGYGR 89  
DB 140 LACLPAGEIILPAGAPCYISGMGR 163  
RESULT 21  
ID 09C052 PRELIMINARY; PRT; 269 AA.  
AC 09C052;  
DT 01-JUN-2001 (TREMBlrel. 17, Created)  
DT 01-JUN-2001 (TREMBlrel. 17, Last sequence update)  
DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)  
DE 2310074F01RIK protein.  
GN ELA3B OR 2310074F01RIK.  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
OX NCBI\_TaxID=10090;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=C57BL/6J; TISSUE=TONGUE;  
RX MEDLINE=21085660; PubMed=11217851;  
RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,  
RA Aikawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,  
RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamana I.,  
RA Saito T., Okazaki Y., Gojobori T., Bono H., Kanukawa T., Saito R.,  
RA Kadoya K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,  
RA Fleischmann W., Gaasterland T., Glass C., King B., Kochia H.,  
RA Kuhl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,  
RA Schirini L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,  
RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,  
RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,  
RA Guernstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,  
RA Guernstein S., Hill D., Hofmann M., Hume D.A., Kamliya M., Lee N.H.,  
RA Lyons P., Matchionni L., Mashima J., Mazzarelli U., Momberts P.,  
RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,  
RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-P.,  
RA Suzuki H., Toyooka K., Wang K.H., Weitz C., Whitaker C., Wilming L.,  
RA Wyshew-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohlsuki S.,  
RA Hayashizaki Y.,  
RT "Functional annotation of a full-length mouse cDNA collection."  
RL Nature 409:685-690(2001).  
CC -! SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE  
TRYPSIN FAMILY.  
DR EMBL; AK010149; BAB26734.1; -;  
DR EMBL; AK009129; BAB26092.1; -;  
DR HSSP; P05805; IFON.  
DR MEROPS; S01.154; -;  
DR MGD; MGI:1915118; Elajb.  
DR InterPro: IPR001314; Chymotrypsin.  
DR InterPro: IPR001254; Ser. protease\_Try.  
DR Pfam; PF00089; trypsin; 1.  
DR PRINTS; PR00722; CHYMOTRYPSIN.  
DR SMART; SM00020; Tryp\_SPC; 1.  
DR PROSITE; PS50240; TRYPSIN\_DOM; 1.  
DR PROSITE; PS00134; TRYPSIN\_HIS; UNKNOWN\_1.  
DR PROSITE; PS00135; TRYPSIN\_SER; 1.  
KW Hydrolase; Serine protease.  
SQ SEQUENCE 269 AA; 28904 MW; C543F76957B2A7CE CRC64;  
  
Query Match 15.5%; Score 85.5; DB 11; Length 269;  
Best Local Similarity 27.4%; Pred. No. 0.087;  
Matches 23; Conservative 14; Mismatches 36; Indels 11; Gaps 1;  
  
QY 6 LKNGDQGIHQPSGVKAPAGYWPSCMSARQRPPIAQLTSGFDIAIVMLAQVNIQSGIR 65  
DB 91 VEEGQEQVPIINAGDLFVHPKNNMVCSC-----GNDIALVKLSRAQLGDAVQ 139  
QY 66 VISLPQSDIPPPGTCVFTVGYGR 89  
DB 140 IACLPPAGELIPNAGAPCYISGWGR 163  
  
RESULT 22  
070500 PRELIMINARY; PRT; 247 AA.  
AC 070500;  
DT 01-AUG-1998 (TREMBLrel. 07, Created)  
DT 01-AUG-1998 (TREMBLrel. 07, Last sequence update)  
DT 01-MAR-2002 (TREMBLrel. 20, Last annotation update)  
DE 01-MAR-2002 (TREMBLrel. 20, Last annotation update)  
DE Vascular chymase.  
GN VCH.  
OS Rattus norvegicus (Rat).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
OX NCBI\_TaxID=10116;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=SPRAGUE-DAWLEY; TISSUE=PULMONARY ARTERY;

RX MEDLINE=21153933; PubMed=11254670;  
RA Guo C., Ju H., Leung D., Massaeli H., Shi M., Rabinovitch M.;  
RT "A novel vascular smooth muscle chymase is upregulated in hypertensive rats";  
RL J. Clin. Invest. 107:703-715(2001).  
CC -! SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE  
TRYPSIN FAMILY.  
DR EMBL; AF063851; AAC16657.1; -;  
DR HSSP; P00770; 3RP2.  
DR MEROPS; S01.095; -;  
DR InterPro: IPR001314; Chymotrypsin.  
DR InterPro: IPR001254; Ser. protease\_Try.  
DR Pfam; PF00089; trypsin; 1.  
DR PRINTS; PR00722; CHYMOTRYPSIN.  
DR SMART; SM00020; Tryp\_SPC; 1.  
DR PROSITE; PS50240; TRYPSIN\_DOM; 1.  
DR PROSITE; PS00134; TRYPSIN\_HIS; UNKNOWN\_1.  
DR PROSITE; PS00135; TRYPSIN\_SER; 1.  
KW Hydrolase; Serine protease.  
SQ SEQUENCE 247 AA; 27278 MW; PF913C1103D6D9F9 CRC64;  
  
Query Match 15.3%; Score 84.5; DB 11; Length 247;  
Best Local Similarity 38.0%; Pred. No. 0.1;  
Matches 19; Conservative 9; Mismatches 21; Indels 1; Gaps 1;  
  
QY 48 DIAIVMLAQVNIQSGIRVLSLPQSDIPPPGTCVFTVGYGRDNDRDP 97  
DB 109 DIMLTKKQVLEPFAVDVPLPSPDIPHPGTLCTWAGWR-TGVKPT 157  
  
RESULT 23  
017030 PRELIMINARY; PRT; 268 AA.  
ID 017030  
AC 017030;  
DT 01-NOV-1996 (TREMBLrel. 01, Created)  
DT 01-FEB-1997 (TREMBLrel. 02, Last sequence update)  
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)  
DE Serine protease.  
OS Anopheles gambiae (African malaria mosquito).  
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
OC Pterygota; Neoptera; Endopterygota; Diptera; Nematocera; Culicidae;  
OC Anopheles.  
OX NCBI\_TaxID=7165;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=G3; TISSUE=MIDGUT;  
RX MEDLINE=97075119; PubMed=8917545;  
RA Dimopoulos G.M., Richman A., della Torre A., Kafatos F.C., Louis C.;  
RT "Identification and characterization of differentially expressed cDNAs  
of the vector mosquito, Anopheles gambiae";  
RL Proc. Natl. Acad. Sci. U.S.A. 93:13066-13071(1996).  
CC -! SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE  
TRYPSIN FAMILY.  
DR EMBL; Z69978; CAA3818.1; -;  
DR HSSP; P00763; IDPO.  
DR InterPro: IPR001314; Chymotrypsin.  
DR InterPro: IPR001254; Ser. protease\_Try.  
DR Pfam; PF00089; trypsin; 1.  
DR PRINTS; PR00722; CHYMOTRYPSIN.  
DR SMART; SM00020; Tryp\_SPC; 1.  
DR PROSITE; PS50240; TRYPSIN\_DOM; 1.  
DR PROSITE; PS00134; TRYPSIN\_HIS; UNKNOWN\_1.  
DR PROSITE; PS00135; TRYPSIN\_SER; 1.  
KW Hydrolase; Protease; Serine protease.  
SQ SEQUENCE 268 AA; 29176 MW; 7BEE6462E5F1D8BE CRC64;  
  
Query Match 15.1%; Score 83.5; DB 5; Length 268;  
Best Local Similarity 27.4%; Pred. No. 0.15;  
Matches 37; Conservative 18; Mismatches 39; Indels 41; Gaps 6;  
  
QY 8 NQDQGIHQPSGVKAPAGY-----MPSCMS-----ARQRRPIA 41  
DB 49 NNDEQDPHFPGGSLIAKFKVLTAGCVPSAISPDGFEAVAGEHDFQYDAGVRRRIA 108





Q9Y1K7  
ID Q9Y1K7 PRELIMINARY; PRT; 365 AA.  
AC Q9Y1K7;  
DT 01-NOV-1999 (TREMBLrel. 12, Created)  
DT 01-NOV-1999 (TREMBLrel. 12, Last sequence update)  
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)  
DE Serine protease 14A.  
GN SPI4A.  
OS Anopheles gambiae (African malaria mosquito).  
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
OC Pterygota; Neoptera; Endopterygota; Diptera; Nematocera; Culicoidae;  
OC Anopheles.  
OX NCBI\_TaxID=7165;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=G3; TISSUE=HEMOLYMPH;  
RX MEDLINE=20110889; PubMed=10646969;  
RA Gorman M.J., Andreeva O.V., Paskewitz S.M.;  
RT "Molecular characterization of five serine protease genes cloned from  
RT Anopheles gambiae hemolymph.";  
RL Insect Biochem. Mol. Biol. 30:35-46(2000).  
CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE  
CC TRYPSIN FAMILY.  
DR EMBL; AF117748; AAC38334.1; -.  
DR HSSP; P20231; IAO.  
DR InterPro; IPR001314; Chymotrypsin.  
DR InterPro; IPR001254; Ser\_protease\_Try.  
DR Pfam; PF00089; trypsin; 1.  
DR PRINTS; PR00722; CHYMOTRYPSIN.  
DR SMART; SM00020; TRYP\_Spc; 1.  
DR PROSITE; PS50240; TRYPsin\_DOM; 1.  
DR PROSITE; PS00134; TRYPsin\_HIS; UNKNOWN\_1.  
DR PROSITE; PS00135; TRYPsin\_SER; 1.  
KW Hydrolase; Protease; Serine protease.  
SQ SEQUENCE 365 AA; 40420 MW; 0C65C5BF6801945 CRC64;

Query Match 14.9%; Score 82; DB 5; Length 365;  
Best Local Similarity 36.1%; Pred. No. 0.32;  
Matches 22; Conservative 9; Mismatches 22; Indels 8; Gaps 3;

QY 48 DIAIYVLAQMVNLQSGIRVSLPQPSDIPP--PGTGVFIYVGRDNDPDRKNGIILK 105  
DB 223 DIALRLDSDVTMNNFVSPCLP-PDDFPPTSGLVNVAVGHGTRGRQ-----HSGIKK 276  
QY 106 K 106  
DB 277 K 277

RESULT 28  
Q13209 PRELIMINARY; PRT; 286 AA.  
AC Q13209;  
DT 01-NOV-1996 (TREMBLrel. 01, Created)  
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)  
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)  
DE Hepatocyte growth factor-like protein homolog (Fragment).  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=LIVER;  
RX MEDLINE=20191171; PubMed=10728827;  
RA Degen S.J.F., McDowell S.A., Waltz S.E., Gould F., Stuart L.A.,  
RA Carritt B.;  
RT "Structure of the human DIF151A locus: a chromosome 1 locus with 97%  
RT identity to the chromosome 3 gene coding for hepatocyte growth factor-  
RT like protein.";  
RL DNA Seq. 8:409-413(1998).  
CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE  
CC TRYPSIN FAMILY.

DR EMBL; U28055; AAC35412.1; -.  
DR HSSP; P00747; SHPG.  
DR MEROPS; S01.975; -.  
DR InterPro; IPR001314; Chymotrypsin.  
DR InterPro; IPR000001; Kringle.  
DR InterPro; IPR001254; Ser\_protease\_Try.  
DR Pfam; PF00051; Kringle; 1.  
DR Pfam; PF00089; trypsin; 1.  
DR PRINTS; PR00722; CHYMOTRYPSIN.  
DR ProDom; PD000395; Kringle; 1.  
DR SMART; SM00130; KR; 1.  
DR SMART; SM00020; TRYP\_Spc; 1.  
DR PROSITE; PS50070; KRINGLE\_2; 1.  
DR PROSITE; PS50240; TRYPsin\_DOM; 1.  
KW Hydrolase; Serine protease.  
FT NON TER 1  
SQ SEQUENCE 286 AA; 31986 MW; 43489B747C8D3F90 CRC64;

Query Match 14.8%; Score 81.5; DB 4; Length 286;  
Best Local Similarity 35.7%; Pred. No. 0.27;  
Matches 20; Conservative 10; Mismatches 23; Indels 3; Gaps 1;

QY 36 QRPVIAQTL---SGFDIAIYVLAQMVNLQSGIRVSLPQPSDIPPCTGVFIYVYG 88  
DB 128 QRPVAKMLCGPSSQLVTLKILERSVTINQKVALICPPENYVVPFGIKCEIAGWG 183

RESULT 29  
Q8RIAG PRELIMINARY; PRT; 331 AA.  
AC Q8RIAG;  
DT 01-JUN-2002 (TREMBLrel. 21, Created)  
DT 01-JUN-2002 (TREMBLrel. 21, Last sequence update)  
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)  
DE RIKEN cDNA 2010001P08 gene.  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
OX NCBI\_TaxID=10090;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=COLON;  
RA Strausberg R.;  
RL Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.  
DR EMBL; BC024903; AAH24903.1; -.  
SQ SEQUENCE 331 AA; 35639 MW; C06F6BF2PA261636 CRC64;

Query Match 14.7%; Score 81; DB 11; Length 331;  
Best Local Similarity 28.6%; Pred. No. 0.37;  
Matches 20; Conservative 13; Mismatches 27; Indels 10; Gaps 1;

QY 29 PSCMSARQRRPIAQT-----SGFDIAIYVLAQMVNLQSGIRVSLPQPSDIPP 78  
DB 116 PEDNEPKELRAVAPRIKHPSYSADSHSGDIALVOLASPIFENDYMLVCPKPDPLDP 175  
QY 79 GTGVFIYVYG 88  
DB 176 GTMCWVTGNG 185

RESULT 30  
O46644 PRELIMINARY; PRT; 266 AA.  
AC O46644;  
DT 01-JUN-1998 (TREMBLrel. 06, Created)  
DT 01-JUN-1998 (TREMBLrel. 06, Last sequence update)  
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)  
DE Elastase I precursor.  
GN ELA1.  
OS Macaca fascicularis (Crab eating macaque) (Cynomolgus monkey).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Cercopithecoidea; Cercopithecinae; Macaca.

OX NCBI\_TaxID=9541;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Kawashima I.;  
RL Submitted (JUG-1995) to the EMBL/GenBank/DDBJ databases.  
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE  
CC TRYPSIN FAMILY.  
DR EMBL; X91400; CA62746.1; -.  
DR HSSP; P00772; 1ELG.  
DR MEROPS; S01.153; -.  
DR InterPro; IPR001314; Chymotrypsin.  
DR InterPro; IPR001254; Ser\_protease\_Try.  
DR Pfam; PF00089; trypsin\_1.  
DR PRINTS; PR00722; CHYMOTRYPSIN.  
DR SMART; SM00020; TRYPSIN\_DOM; 1.  
DR PROSITE; PS50240; TRYPSIN\_DOM; 1.  
DR PROSITE; PS00134; TRYPSIN\_HIS; UNKNOWN\_1.  
DR PROSITE; PS00135; TRYPSIN\_SER; 1.  
KW Hydrolyase; Serine protease; Signal.  
FT SIGNAL 1 16 POTENTIAL.  
FT CHAIN 27 266 POTENTIAL.  
SQ SEQUENCE 266 AA; 28754 MW; 683281B90F53A91C CRC64;  
  
Query Match 14.6%; Score 80.5; DB 6; Length 266;  
Best Local Similarity 35.1%; Pred. No. 0.33;  
Matches 20; Conservative 12; Mismatches 18; Indels 7; Gaps 1;  
  
OY 45 SGPIAIYMLAQMNLQSGIRVLSLPQSDIPPGTGVIYXGRDDNDSDRSKRG 101  
DB 116 AGYIALRLRLAQSIVLNSVQLGVLPQEGAILANDSPCTITGMGR-----TKING 165

## RESULT 31

Q8T3A1 PRELIMINARY; PRT; 470 AA.  
AC Q8T3A1;  
DT 01-JUN-2002 (TREMBlrel. 21, Created)  
DT 01-JUN-2002 (TREMBlrel. 21, Last sequence update)  
DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)  
DE Putative coagulation serine protease.  
GN SP3.  
OS Ciona intestinalis.  
OC Eukaryota; Metazoa; Chordata; Urochordata; Ascidiacea; Enterogona;  
OC Plebobranchia; Clonidae; Ciona.  
OX NCBI\_TaxID=7719;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=HEPATOPANCREAS;  
RA Hammond J.A., Nakao M., Yano T., Kemp G.D., Smith V.J.;  
RT "Complete cDNA sequence of a serine protease from Ciona  
RT intestinalis.";  
RL Submitted (FEB-2002) to the EMBL/GenBank/DDBJ databases.  
DR EMBL; AJ31685; CAD24308.1; -.  
KW Protease.  
SQ SEQUENCE 470 AA; 52172 MW; 284259711A38B1E7 CRC64;  
  
Query Match 14.6%; Score 80.5; DB 5; Length 470;  
Best Local Similarity 33.3%; Pred. No. 0.65;  
Matches 22; Conservative 12; Mismatches 23; Indels 9; Gaps 3;  
  
OY 35 RQRPIAQL--SGF-----DIAIVMLAQMNLQSGIRVLSLPQSDIPPGTGVIYXGR 86  
DB 208 RQRREVEIYIHFGFTAYLVNDVALIKLSRPVENDITIPICLP-CGETPSGDKCWYTG 266  
OY 87 YGRDNN 92  
DB 267 FGRTEN 272  
  
RESULT 32  
Q924U6 PRELIMINARY; PRT; 200 AA.  
ID Q924U6  
AC Q924U6;

DT 01-DEC-2001 (TREMBlrel. 19, Created)  
DT 01-DEC-2001 (TREMBlrel. 19, Last sequence update)  
DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)  
DE Serine protease-like 1.  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
OX NCBI\_TaxID=10090;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Matsui H., Takahashi T.;  
RT "Mus musculus mRNA for serine protease-like gene spliced variant-1.";  
RL Submitted (AUG-2000) to the EMBL/GenBank/DDBJ databases.  
DR EMBL; AB047758; BAB63919.1; -.  
DR MEROPS; S01.106; -.  
DR InterPro; IPR001254; Ser\_protease\_Try.  
DR Pfam; PF00089; trypsin\_1.  
DR PROSITE; PS50240; TRYPSIN\_DOM; 1.  
DR PROSITE; PS00135; TRYPSIN\_SER; UNKNOWN\_1.  
KW Hydrolyase; Protease; Serine protease.  
SQ SEQUENCE 200 AA; 22135 MW; 972959E4686BF603 CRC64;  
  
Query Match 14.5%; Score 80; DB 11; Length 200;  
Best Local Similarity 38.1%; Pred. No. 0.27;  
Matches 16; Conservative 11; Mismatches 15; Indels 0; Gaps 0;  
  
OY 48 DIAIVMLAQMNLQSGIRVLSLPQSDIPPGTGVIYXGR 89  
DB 34 DIAIVLAFVNVSVNIQPCIPKSFVLPQITLCWYGMGR 75

## RESULT 33

Q9XYX9 PRELIMINARY; PRT; 248 AA.  
AC Q9XYX9;  
DT 01-NOV-1999 (TREMBlrel. 12, Created)  
DT 01-NOV-1999 (TREMBlrel. 12, Last sequence update)  
DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)  
DE Trypsinogen Rdot1.  
OS Rhyzopertha dominica (lesser grain borer).  
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
OC Pterygota; Neoptera; Endopterygota; Coleoptera; Polyphaga;  
OC Bostrichiformia; Bostrichidae; Rhyzopertha.  
OX NCBI\_TaxID=92692;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=20078164; PubMed=10612040;  
RA Zhu Y.C., Baker J.E.;  
RT "Characterization of midgut trypsin-like enzymes and three trypsinogen  
RT cDNAs from the lesser grain borer, Rhyzopertha dominica (Coleoptera:  
RT Bostrichidae).";  
RL Insect Biochem. Mol. Biol. 29:1053-1063(1999).  
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE  
CC TRYPSIN FAMILY.  
DR EMBL; AF130840; AAD31267.1; -.  
DR HSSP; P00763; IDPO.  
DR MEROPS; S01.113; -.  
DR InterPro; IPR001314; Chymotrypsin.  
DR InterPro; IPR003006; IG\_MHC.  
DR InterPro; IPR001254; Ser\_protease\_Try.  
DR Pfam; PF00089; trypsin\_1.  
DR PRINTS; PR00722; CHYMOTRYPSIN.  
DR SMART; SM00020; TRYPSIN\_DOM; 1.  
DR PROSITE; PS00290; IG\_MHC; UNKNOWN\_1.  
DR PROSITE; PS50240; TRYPSIN\_DOM; 1.  
DR PROSITE; PS00135; TRYPSIN\_SER; 1.  
KW Hydrolyase; Serine protease.  
SQ SEQUENCE 248 AA; 26067 MW; 1AB88C3952E3863E CRC64;  
  
Query Match 14.5%; Score 80; DB 5; Length 248;  
Best Local Similarity 24.6%; Pred. No. 0.35;  
Matches 14; Conservative 15; Mismatches 28; Indels 0; Gaps 0;

[illegible]



## RESULT 40

```

097366 PRELIMINARY; PRT; 365 AA.
ID 097366
AC 097366;
DT 01-MAY-1999 (TREMBlrel. 10, Created)
DT 01-MAY-1999 (TREMBlrel. 10, last sequence update)
DT 01-MAR-2002 (TREMBlrel. 20, last annotation update)
DE Pro-phenoloxidase activating enzyme-1 precursor.
OS Holotrichia diomphalia.
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Coleoptera; Polyphaga;
OC Scarabaeiformia; Scarabaeidae; Melolonthinae; Holotrichia.
OX NCBI_TaxID=33394;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=99053686; PubMed=9839951;
RA Lee S.Y., Cho M.Y., Hyun J.H., Lee K.M., Homma K., Natori S.,
RA Kawabata S., Iwanaga S., Lee B.L.;
RT "Molecular cloning of cDNA for pro-phenol oxidase-activating factor 1,
RT a serine protease that induced by lipopolysaccharide or 1,3-beta-
RT glucan in coleopteran insect, Holotrichia diomphalia larvae.";
RL Eur. J. Biochem. 257:615-621(1998)
CC -! SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE
CC TRYPSIN FAMILY.
DR EMBL; AB013088; BAA34642.1; -.
DR HSSP; P00763; IDPO.
DR MEROPS; S01.204; -.
DR InterPro; IPR001314; Chymotrypsin.
DR InterPro; IPR001254; Ser_protease TRY.
DR Pfam; PF00089; trypsin; 1.
DR PRINTS; PR00722; CHYMOTRYPSIN.
DR SMART; SM00020; Tryp_SPC; 1.
DR PROSITE; PS50240; TRYPSIN_DOM; 1.
DR PROSITE; PS00134; TRYPSIN_HIS; UNKNOWN_1.
DR PROSITE; PS00135; TRYPSIN_SER; 1.
KW Hydrolyase; Serine protease; Signal.
FT SIGNAL 1 21 POTENTIAL.
FT SIGNAL 22 109 POTENTIAL.
FT CHAIN 110 365 POTENTIAL.
SQ SEQUENCE 365 AA; 40194 MW; EC69F6093923F4C5 CRC64;

Query Match 14.1%; Score 78; DB 5; Length 365;
Best Local Similarity 36.4%; Pred. No. 0.94;
Matches 16; Conservative 11; Mismatches 17; Indels 0; Gaps 0;

```

CY 48 DIAIVMLAOMVNLQSGIRIVISLPQPSDIPPPGTGVFIYVGRDD 91

DB 220 DIALIRLNQVEFTYIRPVCLPQNEEVQVGORLTVGVGRTE 263

Search completed: April 9, 2003, 12:57:59  
Job time : 96.8361 secs

GenCore version 5.1.4 p5 4578  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - protein search, using frame\_plus\_n2p model

Run on: April 9, 2003, 13:27:38 / Search time 24 Seconds

(without alignments)  
3621.471 Million cell updates/sec

Title: US-10-020-441-1

Perfect score: 2686  
Sequence: 1 gtgtcaaccggtgtacacg.....ctgtacgtccaatattga 1477

Scoring table:

BLOSUM62  
Xgapop 10.0, Xgapext 0.5  
Fgapop 6.0, Fgapext 7.0  
Delop 6.0, Delext 7.0

Searched: 262574 seqs, 29422922 residues

Total number of hits satisfying chosen parameters: 525148

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:  
-MODEL=frame\_n2p.model -DEV=xlp  
-O=/cgn2\_1/USPTO.epool/US10020441/runat\_09042003\_121212\_1379/app\_query.fasta\_1.1671  
-DB=Issued\_Patents\_AA -QPMT=fastan -SUFFIX=rai -MINMATCH=0.1 -LOOPCL=0  
-LOOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=Dlosum62 -TRANS=human40.cdi  
-LIST=45 -DOCALIGN=200 -THR SCORE=PCT -THR MAX=100 -THR MIN=0 -ALIGN=40  
-MODE=LOCAL -OUTFMT=pcio -NORM=ext -HEAPSIZE=500 -MITLEN=0 -MAXLEN=2000000000  
-USRR=US10020441 @CGN 1.1 28 @runat\_09042003\_121212\_1379 -NCPU=6 -ICPU=3  
-NO\_XLPHY -NO\_MMAR -LARGEOVERY -NEG SCORES=0 -WAIT -LONGLOG -DEV TIMEOUT=120  
-WARN TIMEOUT=30 -THRAD=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7  
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

Issued Patents AA:\*  
1: /cgn2\_6/ptodata/1/iaa/5A.COMB.pep:\*  
2: /cgn2\_6/ptodata/1/iaa/5B.COMB.pep:\*  
3: /cgn2\_6/ptodata/1/iaa/6A.COMB.pep:\*  
4: /cgn2\_6/ptodata/1/iaa/6B.COMB.pep:\*  
5: /cgn2\_6/ptodata/1/iaa/PCTUS.COMB.pep:\*  
6: /cgn2\_6/ptodata/1/iaa/backfiles1.pep:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	130.5	4.9	190	2	US-08-845-998-4
2	130.5	4.9	190	4	US-09-205-537-4
3	130.5	4.9	190	4	US-09-430-854-4
4	126.5	4.7	190	4	US-08-845-998-6
5	126.5	4.7	190	4	US-09-206-537-6
6	126.5	4.7	190	4	US-09-430-854-6
7	123	4.6	273	2	US-09-016-366A-19
8	123	4.6	273	2	US-08-978-404B-14
9	119	4.4	245	4	US-08-944-483-69
10	119	4.4	245	4	US-09-016-366A-17
11	119	4.4	275	2	US-08-978-404B-12
12	117	4.4	245	4	US-09-079-970A-6

13	117	4.4	249	4	US-09-079-970A-5	Sequence 5, Appli
14	117	4.4	267	2	US-09-016-366A-23	Sequence 23, Appli
15	117	4.4	267	2	US-08-978-404B-18	Sequence 18, Appli
16	117	4.4	270	2	US-08-978-404B-8	Sequence 8, Appli
17	117	4.4	274	2	US-09-016-366A-21	Sequence 21, Appli
18	117	4.4	274	2	US-08-978-404B-16	Sequence 16, Appli
19	116.5	4.3	242	4	US-08-944-483-57	Sequence 57, Appli
20	114.5	4.3	242	4	US-08-944-483-57	Sequence 58, Appli
21	113.5	4.3	1721	3	US-08-928-361B-6	Sequence 5, Appli
22	113.5	4.3	1721	3	US-08-928-361B-6	Sequence 6, Appli
23	111	4.1	276	2	US-09-016-366A-15	Sequence 15, Appli
24	111	4.1	276	2	US-08-978-404B-21	Sequence 21, Appli
25	107.5	4.0	1837	3	US-08-928-361B-5	Sequence 5, Appli
26	107	4.0	273	2	US-08-978-404B-6	Sequence 6, Appli
27	107	4.0	306	1	US-08-414-926A-16	Sequence 16, Appli
28	107	4.0	306	2	US-08-926-922-16	Sequence 16, Appli
29	107	4.0	306	3	US-09-253-682-16	Sequence 16, Appli
30	107	4.0	306	4	US-09-527-657-16	Sequence 16, Appli
31	104.5	3.9	249	3	US-08-700-651-15	Sequence 15, Appli
32	104.5	3.9	249	3	US-08-928-361B-20	Sequence 20, Appli
33	102.5	3.8	222	4	US-09-605-785-479	Sequence 479, App
34	102.5	3.8	222	4	US-09-439-113-479	Sequence 479, App
35	101.5	3.8	216	3	US-08-928-361B-27	Sequence 27, Appli
36	100.5	3.8	216	3	US-08-928-361B-5	Sequence 8, Appli
37	100	3.7	274	2	US-08-978-404B-5	Sequence 5, Appli
38	99	3.7	273	2	US-08-978-404B-3	Sequence 3, Appli
39	99	3.7	290	4	US-09-386-653A-7	Sequence 7, Appli
40	99	3.7	315	4	US-09-386-653A-9	Sequence 9, Appli
41	98.5	3.7	224	1	US-08-278-081-12	Sequence 12, Appli
42	98.5	3.7	224	1	US-08-483-859-12	Sequence 12, Appli
43	98.5	3.7	224	1	US-08-472-173-12	Sequence 12, Appli
44	98.5	3.7	224	2	US-08-487-167-12	Sequence 12, Appli
45	98.5	3.7	224	2	US-08-482-816-12	Sequence 12, Appli

## ALIGNMENTS

RESULT 1  
US-08-845-998-4  
Sequence 4, Application US/08845998  
Patent No. 5879892  
GENERAL INFORMATION:  
APPLICANT: Van Baren, Nicolas  
APPLICANT: Coulie, Pierre G.  
APPLICANT: De Smet, Charles  
APPLICANT: Lucas, Sophie  
APPLICANT: Boon, Thierry  
TITLE OF INVENTION: LEUKEMIA ASSOCIATED GENES  
NUMBER OF SEQUENCES: 16  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Wolf, Greenfield & Sacks, P.C.  
STREET: 600 Atlantic Avenue  
CITY: Boston  
STATE: MA  
COUNTRY: US  
ZIP: 02210  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/845,998  
FILING DATE:  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Van Amsterdam, John R.  
REGISTRATION NUMBER: 40,212  
REFERENCE/DOCKET NUMBER: L0461/7008  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (617)720-3500  
TELEFAX: (617)720-2441  
INFORMATION FOR SEQ ID NO: 4:





Db 158 ThrGlyTPrGlyAspValAspAsn-----AsnGlyGly----- 168  
 Oy 622 AAGAAAGTGAAGTGTGG 639  
 Db 169 -----CysTrp 170

RESULT 3

US-09-430-854-4  
 Sequence 4, Application US/09430854  
 Patent No. 6271019

GENERAL INFORMATION:

APPLICANT: Van Baren, Nicolas  
 APPLICANT: Coulie, Pierre G.  
 APPLICANT: De Smet, Charles  
 APPLICANT: Lucas, Sophie  
 APPLICANT: Boon, Thierry  
 TITLE OF INVENTION: LEUKEMIA ASSOCIATED GENES  
 NUMBER OF SEQUENCES: 16  
 CORRESPONDENCE ADDRESSES:  
 ADDRESSEE: Wolf, Greenfield & Sacks, P.C.  
 STREET: 600 Atlantic Avenue  
 CITY: Boston  
 STATE: MA  
 COUNTRY: US

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk  
 COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: PatentIn Release #1.0, Version #1.25  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/09/430,854  
 FILING DATE:  
 CLASSIFICATION:  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: 08/845,998  
 FILING DATE:  
 ATTORNEY/AGENT INFORMATION:  
 NAME: Van Amsterdam, John R.  
 REGISTRATION NUMBER: 40,212  
 REFERENCE/DOCKET NUMBER: L0461/7008  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: (617)720-3500  
 TELEFAX: (617)720-2441  
 INFORMATION FOR SEQ ID NO: 4:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 190 amino acids  
 TYPE: amino acid  
 TOPOLOGY: linear  
 MOLECULE TYPE: protein  
 US-09-430-854-4

Alignment Scores:

Pred. No.: 1.96e-05 Length: 190  
 Score: 130.50 Matches: 42  
 Percent Similarity: 34.95% Conservative: 23  
 Best Local Similarity: 22.58% Mismatches: 44  
 Query Match: 4.86% Indels: 77  
 DB: 4 Gaps: 5

US-10-020-441-1 (1-1477) x US-09-430-854-4 (1-190)

Oy 91 TGTACAGTTTACATGTTCTCAACGACGATCTCACAGTGTGATTTGTT----- 144  
 Db 59 CysGlyGlySerLeuIleHisProGlnTrpValLeuThrAlaAlaHisCysValGluPro 78  
 Oy 145 ---TGCTCACCATTGCGCAGTATTCGGGTAAGAGATCGATGAACAACATTTGCTCAA 201  
 Db 79 AspileysaapleuAlaAlaLeuArgValGlnLeuArgGluGlnHisIleuYrTrpGln 98  
 Oy 202 TGTAAATCATTTATTCACAGACAGTGTGTTGGTGGTGGCTTTGTGTTGTGTATGC 261

Db 99 AspGln----- 100  
 Oy 262 AGTGGTTTGATAGATTTTATATGCTATATATCTTATTTCAAGTTTCAATTTCTC 321  
 Db 101 -----LeuLeuProValSerArgIleIleValHisProGlnPheYr 114  
 Oy 322 AACTGAGGAATGGCGACCAACAGGATCCATGCACCAACCGTCTGGATTAAGTGGCA 381  
 Db 115 IleIleGln----- 117  
 Oy 382 CCAGATCATGCCCTCTTATATGTGGCAGACAGAGAGAACCAATGCAACACTC 441  
 Db 117 ----- 117  
 Oy 442 AGTGATTCGATATTCATTTGATATGCTGGCTCAAAATGTCATTCACAGTGAATC 501  
 Db 118 ThrGlyAlaAspIleAlaLeuGluLeuGluGluProValAsnIleSerHisIle 137  
 Oy 502 AAGATGATCAGTCTGCCACACCATCGGATATCCCGCACCTGGAAGTGTGTTTCATT 561  
 Db 138 HisThrValThrIleuProAlaSerGluThrPheProGlyMetProCysTrpVal 157  
 Oy 562 GTTGTTATGGAAGGATGATTAACGACCGTATCCGTCACGTAAGATGTGGAATATTG 621  
 Db 158 ThrGlyTPrGlyAspValAspAsn-----AsnGlyGly----- 168  
 Oy 622 AAGAAAGTGAAGTGTGG 639  
 Db 169 -----CysTrp 170

RESULT 4

US-08-845-998-6  
 Sequence 6, Application US/08845998  
 Patent No. 5879892

GENERAL INFORMATION:

APPLICANT: Van Baren, Nicolas  
 APPLICANT: Coulie, Pierre G.  
 APPLICANT: De Smet, Charles  
 APPLICANT: Lucas, Sophie  
 APPLICANT: Boon, Thierry  
 TITLE OF INVENTION: LEUKEMIA ASSOCIATED GENES  
 NUMBER OF SEQUENCES: 16  
 CORRESPONDENCE ADDRESSES:  
 ADDRESSEE: Wolf, Greenfield & Sacks, P.C.  
 STREET: 600 Atlantic Avenue  
 CITY: Boston  
 STATE: MA  
 COUNTRY: US

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk  
 COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: PatentIn Release #1.0, Version #1.25  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/08/845,998  
 FILING DATE:  
 CLASSIFICATION: 435  
 ATTORNEY/AGENT INFORMATION:  
 NAME: Van Amsterdam, John R.  
 REGISTRATION NUMBER: 40,212  
 REFERENCE/DOCKET NUMBER: L0461/7008  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: (617)720-3500  
 TELEFAX: (617)720-2441  
 INFORMATION FOR SEQ ID NO: 6:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 190 amino acids  
 TYPE: amino acid  
 TOPOLOGY: linear  
 MOLECULE TYPE: protein  
 US-08-845-998-6

Alignment Scores: 190  
Pred. No.: 5.18e-05  
Score: 126.50  
Percent Similarity: 34.41%  
Best Local Similarity: 22.04%  
Query Match: 4.71%  
Matches: 41  
Conservative: 23  
Mismatches: 45  
Indels: 77  
Gaps: 5

US-10-020-441-1 (1-1477) x US-08-845-998-6 (1-190)

```
QY 91 TGTACAGSTTCCTAGTCTCAACGAGACGATACACAGCTGTCATTGCTT----- 144
DB 59 CysglglySerPheIleHisProGlnTrpValLeuThrAlaIleHisCysValGluPro 78
QY 145 ---TGCACCACTTGCACAGTATTCGGGTAGAGATCGACTGAACACATTTGCTCAA 201
DB 79 AspleIeYsAspleuAlaIleuArGValGlnLeuArGValGlnHisIleuTyTrGln 98
QY 202 TGTAAATCATGATTTCACAGACAGATGTTGGGTGGCGTGGCTTTGTGTGTATGC 261
DB 99 AspGln----- 100
QY 262 AGTGGGTTGCATAGATTTTAAATGCTATATCTGCTTATTTACAGTTTCATTCTC 321
DB 101 -----LeuLeuProValIleSerArGilleIleValHisProGlnPheTyR 114
QY 322 ACACAGAGGAATGGCGACCAACAGGATTCATCACCAACCGTGTGAAGTTAAGTGCA 381
DB 115 IleIleGln----- 117
QY 382 CCAGATACATGCGCTTGTGTATGTCGGCAGACAGAGAGACCAATTCGCACAGACTC 441
DB 117 ----- 117
QY 442 AGTGATTCGATATTCGAATTTGTAATGCTGCTCAAAATGTCATCTACAGATGAATC 501
DB 118 ThrGlyAlaAspleIleAlaLeuLeuGlnLeuGlnProValAsnIleSerHisIle 137
QY 502 AGAGATGATCAGTCTGCACAGCCATCGATATCCGCCACCTGGAACCTGGTTCATT 561
DB 138 HisThrValThrLeuProProAlaSerGluThrPheProGlyMetProCysTrpVal 157
QY 562 GTTGTTATGAGAGGATGATTAACGACCGTGATCCGTACGTAAGAAGTGGAAATATG 621
DB 158 ThrGlyTrpGlyAspValAspAsn-----AsnGlyGly----- 168
QY 622 AAGAAAGTGAAGTTGTTGG 639
DB 169 -----CysTrp 170
```

## RESULT 5

US-09-206-537-6  
Sequence 6, Application US/09206537  
Patent No. 6130052

## GENERAL INFORMATION:

APPLICANT: Van Baren, Nicolas  
APPLICANT: Coulie, Pierre G.  
APPLICANT: De Smet, Charles  
APPLICANT: Lucas, Sophie  
APPLICANT: Boon, Thierry  
TITLE OF INVENTION: LEUKEMIA ASSOCIATED GENES  
NUMBER OF SEQUENCES: 16  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Wolf, Greenfield & Sacks, P.C.  
STREET: 600 Atlantic Avenue  
CITY: Boston  
STATE: MA  
COUNTRY: US  
ZIP: 02210

## COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/206,537  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/845,998  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: Van Amsterdam, John R.  
REGISTRATION NUMBER: 40,212  
REFERENCE/DOCKET NUMBER: 10461/7008  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (617) 720-3500  
TELEFAX: (617) 720-2441  
INFORMATION FOR SEO ID NO: 6:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 190 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-09-206-537-6

Alignment Scores: 190  
Pred. No.: 5.18e-05  
Score: 126.50  
Percent Similarity: 34.41%  
Best Local Similarity: 22.04%  
Query Match: 4.71%  
Matches: 41  
Conservative: 23  
Mismatches: 45  
Indels: 77  
Gaps: 5

US-10-020-441-1 (1-1477) x US-09-206-537-6 (1-190)

```
QY 91 TGTACAGSTTCCTAGTCTCAACGAGACGATACACAGCTGTCATTGCTT----- 144
DB 59 CysglglySerPheIleHisProGlnTrpValLeuThrAlaIleHisCysValGluPro 78
QY 145 ---TGCACCACTTGCACAGTATTCGGGTAGAGATCGACTGAACACATTTGCTCAA 201
DB 79 AspleIeYsAspleuAlaIleuArGValGlnLeuArGValGlnHisIleuTyTrGln 98
QY 202 TGTAAATCATGATTTCACAGACAGATGTTGGGTGGCGTGGCTTTGTGTGTATGC 261
DB 99 AspGln----- 100
QY 262 AGTGGGTTGCATAGATTTTAAATGCTATATCTGCTTATTTACAGTTTCATTCTC 321
DB 101 -----LeuLeuProValIleSerArGilleIleValHisProGlnPheTyR 114
QY 322 ACACAGAGGAATGGCGACCAACAGGATTCATCACCAACCGTGTGAAGTTAAGTGCA 381
DB 115 IleIleGln----- 117
QY 382 CCAGATACATGCGCTTGTGTATGTCGGCAGACAGAGAGACCAATTCGCACAGACTC 441
DB 117 ----- 117
QY 442 AGTGATTCGATATTCGAATTTGTAATGCTGCTCAAAATGTCATCTACAGATGAATC 501
DB 118 ThrGlyAlaAspleIleAlaLeuLeuGlnLeuGlnProValAsnIleSerHisIle 137
QY 502 AGAGATGATCAGTCTGCACAGCCATCGATATCCGCCACCTGGAACCTGGTTCATT 561
DB 138 HisThrValThrLeuProProAlaSerGluThrPheProGlyMetProCysTrpVal 157
QY 562 GTTGTTATGAGAGGATGATTAACGACCGTGATCCGTACGTAAGAAGTGGAAATATG 621
DB 158 ThrGlyTrpGlyAspValAspAsn-----AsnGlyGly----- 168
QY 622 AAGAAAGTGAAGTTGTTGG 639
DB 169 -----CysTrp 170
```

RESULT 6  
US-09-430-854-6  
; Sequence 6, Application US/09430854  
; Patent No. 6271019  
; GENERAL INFORMATION:  
; APPLICANT: Van Baren, Nicolas  
; APPLICANT: Coulie, Pierre G.  
; APPLICANT: De Smet, Charles  
; APPLICANT: Lucas, Sophie  
; APPLICANT: Boon, Thierry  
; TITLE OF INVENTION: LEUKEMIA ASSOCIATED GENES  
; NUMBER OF SEQUENCES: 16  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Wolf, Greenfield & Sacks, P.C.  
; STREET: 600 Atlantic Avenue  
; CITY: Boston  
; STATE: MA  
; COUNTRY: US  
; ZIP: 02210  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/430,854  
; FILING DATE:  
; CLASSIFICATION:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 09/845,998  
; FILING DATE:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Van Amerdam, John R.  
; REGISTRATION NUMBER: 40,212  
; REFERENCE/DOCKET NUMBER: L0461/7008  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (617)720-3500  
; TELEFAX: (617)720-2441  
; INFORMATION FOR SEQ ID NO: 6:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 190 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; US-09-430-854-6  
  
Alignment Scores:  
Pred. No.: 5,18e-05 Length: 190  
Score: 126.50 Matches: 41  
Percent Similarity: 34.41% Conservative: 23  
Best Local Similarity: 22.04% Mismatches: 45  
Query Match: 4.71% Indels: 77  
DB: 4 Gaps: 5  
  
US-10-020-441-1 (1-1477) x US-09-430-854-6 (1-190)  
QY 91 TGTACAGTCTCACTGCTCAACGAGAGACTCAAGCTGCTCATTTGTT----- 144  
DB 59 Cysglglyserphenilserprogintrpvalleuthralalhisecysvalguipro 78  
QY 145 ---TGTCCACCATTCGACGATTCGGGTAGAGATTCAGTAAACACATTTGCTCA 201  
DB 79 Asplelysaapleuhalalaleuhyralglnleuargluginhisleuylrryrgln 98  
QY 202 TGTAAATCAATTGATTTCACAGACAGAGCTTTGGCTGCGCTTGTGTGTATGTC 261  
DB 99 Asggin----- 100  
QY 262 AGTGGTTTGCATAGATTTTAATGCTAATACCTGCTTATTTCAAGTTTCATTTC 321  
DB 101 -----LeuLeuProValSerArgIleIleValHisProGlnPheTyr 114  
QY 322 ACACTGAGGAATGGCGACCAACAAGGATTCATCAACCAACCGCTGAGATTAGTGCA 381

DB 115 Ilelelgln----- 117  
QY 382 CCAGATACATGACCTCTTGTATGTGCGACAGACAGAGACCAATGCGACACACTC 441  
DB 117 ----- 117  
QY 442 AGTGGATTCATATTGCAATTTGTAATGCTGCTCAATATGTCATTCACAGTGAATC 501  
DB 118 ThrGlyAlaaplelealaleuLeuenglueuglnuproValAsnIleSerSerHisIle 137  
QY 502 AGAGTATCATCTGCTCCACAGCATTCGATATCCGCCACTGGAACGTGTTCATT 561  
DB 138 HisThrValThrleuProProAlaSerGlnThrPheProGlyMetProCysTyrVal 157  
QY 562 GTTGGTTATGGAAGGATTAAGACCGGTATCCGTCACGTAAGATGTGGAATATTG 621  
DB 158 ThrGlyTyrGlyAspValAspAsn-----AsnGlyGly----- 168  
QY 622 AAGAAAGTGAAGTTGTGG 639  
DB 169 -----CysTyr 170  
  
RESULT 7  
US-09-016-366A-19  
; Sequence 19, Application US/09016366A  
; Patent No. 5955431  
; GENERAL INFORMATION:  
; APPLICANT: Stevens, Richard L.  
; APPLICANT: Huang, Chitu  
; TITLE OF INVENTION: MAST CELL PROTEASE PEPTIDE  
; NUMBER OF SEQUENCES: 65  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Wolf, Greenfield & Sacks, P.C.  
; STREET: 600 Atlantic Avenue  
; CITY: Boston  
; STATE: MA  
; COUNTRY: U.S.A.  
; ZIP: 02210-2211  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette  
; COMPUTER: IBM Compatible  
; OPERATING SYSTEM: DOS  
; SOFTWARE: FastSeq for Windows Version 2.0  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/016,366A  
; FILING DATE: January 30, 1998  
; CLASSIFICATION: 530  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 60/037,090  
; FILING DATE: 05-FEB-1997  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Plummer, Elizabeth R.  
; REGISTRATION NUMBER: 36,637  
; REFERENCE/DOCKET NUMBER: B0801/7093  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 617-720-3500  
; TELEFAX: 617-720-2441  
; TELEX:  
; INFORMATION FOR SEQ ID NO: 19:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 273 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; US-09-016-366A-19  
  
Alignment Scores:  
Pred. No.: 0.000142 Length: 273  
Score: 123.00 Matches: 38  
Percent Similarity: 35.50% Conservative: 22  
Best Local Similarity: 22.49% Mismatches: 45

Query Match: 4.58% Indels: 64  
DB: 2 Gaps: 3

US-10-020-441-1 (1-1477) x US-09-016-366A-19 (1-273)

```

QY 91 TGTACAGSTTCCTAGTCTCAACAGACAGACTACTCAAGCTGGTCATTGTGTT----- 144
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 57 CysgIyglYserLeuIleHISProGInTrPValleuThraIalalHIScyValGIyPro 76
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 145 ---TGTCAACCATTCAGTGGTGGGTGAAGATCGCATCAACACATTTGCTGCA 201
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 77 AspValILysAspLeuAlalaleuLysGValGInLeuLysGluGlnHisLeuYrTyrgIn 96
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 202 TGTATTCGATTGATTTCACAGACAGAGCTGTGGCTGGCTGTGGTTGTGTATGC 261
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 97 AspGln----- 98
QY 262 AGTGGGTTTGCAAGATTTTAATGCTATATACCTTGCTTATTTACAGTTTCATTCTC 321
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 99 -----LeuLeuProValSerArgIleIleValHisProGInPheTy 112
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 332 ACACGTGAGGAATGGGACCAACAAGAGCATCCATCAACACGCTTGAGTTAAGTGGA 381
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 113 ThrAlaGln----- 115
QY 382 CCAGATACATGCCCTCTGTATGTGCGACAGACAGAGAGACAATCGCACAGACATC 441
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 115 ----- 115
QY 442 ACTGATTCGATATTCGATTTGTAATGCTGCTCAAAATGCTCACTTACAGATGGAATC 501
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 116 IlegIyAlaAspIleAlaleuLeuGlnLeuGluProValAspValSerHisVal 135
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 502 AGAGTACAGCTGCCACAGCATCGGATATCCGGCCACCTGGAACCTGTTTCATT 561
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 136 HisThrValThrLeuProProAlaSerGluThrPheProGlyMecProCysTrpVal 155
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 562 GTTGGTTATGGAAGGATGATTAACGAC 588
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 156 ThrGlyTrpGlyAspValAspAspAsp 164
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 8
US-08-978-404B-14
; Sequence 14, Application US/08978404B
; Patent No. 5968782
; GENERAL INFORMATION:
; APPLICANT: Stevens, Richard L.
; TITLE OF INVENTION: MAST CELL PROTEASE THAT CLEAVES
; TITLE OF INVENTION: FIBRINOGEN
; NUMBER OF SEQUENCES: 74
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Wolf, Greenfield & Sacks, P.C.
; STREET: 600 Atlantic Avenue
; CITY: Boston
; STATE: MA
; COUNTRY: U.S.A.
; ZIP: 02210-2211
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FASTSEQ for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/978,404B
; FILING DATE: 25-NOV-97
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 60/032,354
; FILING DATE: 04-DEC-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Plumer, Elizabeth R.
; REGISTRATION NUMBER: 36,637
; REFERENCE/DOCKET NUMBER: B0801/7090

```

# TELECOMMUNICATION INFORMATION:

TELEPHONE: 617-720-3500  
TELEFAX: 617-720-2441

TELEX: INFORMATION FOR SEQ ID NO: 14:

SEQUENCE CHARACTERISTICS:

LENGTH: 273 amino acids

TYPE: amino acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: No. 5968782e

US-08-978-404B-14

Alignment Scores:

Pred. No.: 0.000142 Length: 273

Score: 123.00 Matches: 38

Percent Similarity: 35.50% Conservative: 22

Best Local Similarity: 22.49% Mismatches: 45

Query Match: 4.58% Indels: 64

DB: 2 Gaps: 3

US-10-020-441-1 (1-1477) x US-08-978-404B-14 (1-273)

```

QY 91 TGTACAGSTTCCTAGTCTCAACAGACAGACTACTCAAGCTGGTCATTGTGTT----- 144
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 57 CysgIyglYserLeuIleHISProGInTrPValleuThraIalalHIScyValGIyPro 76
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 145 ---TGTCAACCATTCAGTGGTGGGTGAAGATCGCATCAACACATTTGCTGCA 201
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 77 AspValILysAspLeuAlalaleuLysGValGInLeuLysGluGlnHisLeuYrTyrgIn 96
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 202 TGTATTCGATTGATTTCACAGACAGAGCTGTGGCTGGCTGTGGTTGTGTATGC 261
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 97 AspGln----- 98
QY 262 AGTGGGTTTGCAAGATTTTAATGCTATATACCTTGCTTATTTACAGTTTCATTCTC 321
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 99 -----LeuLeuProValSerArgIleIleValHisProGInPheTy 112
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 332 ACACGTGAGGAATGGGACCAACAAGAGCATCCATCAACACGCTTGAGTTAAGTGGA 381
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 113 ThrAlaGln----- 115
QY 382 CCAGATACATGCCCTCTGTATGTGCGACAGACAGAGAGACAATCGCACAGACATC 441
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 115 ----- 115
QY 442 ACTGATTCGATATTCGATTTGTAATGCTGCTCAAAATGCTCACTTACAGATGGAATC 501
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 116 IlegIyAlaAspIleAlaleuLeuGlnLeuGluProValAspValSerHisVal 135
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 502 AGAGTACAGCTGCCACAGCATCGGATATCCGGCCACCTGGAACCTGTTTCATT 561
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 136 HisThrValThrLeuProProAlaSerGluThrPheProGlyMecProCysTrpVal 155
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 562 GTTGGTTATGGAAGGATGATTAACGAC 588
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 156 ThrGlyTrpGlyAspValAspAspAsp 164
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 9
US-08-944-483-69
; Sequence 69, Application US/08944483
; Patent No. 6232456
; GENERAL INFORMATION:
; APPLICANT: COHEN, MAURICE
; APPLICANT: COLPITTS, TRACEY L.
; APPLICANT: FRIEDMAN, PAULA N.
; APPLICANT: GRANADOS, EDWARD N.
; APPLICANT: KLAAS, MICHAEL R.
; APPLICANT: RUSSELL, JOHN C.
; APPLICANT: STEWART, KENT D.
; APPLICANT: STROPE, STEVEN D.
; TITLE OF INVENTION: NOVEL SERINE PROTEASE REAGENTS

```

[illegible]

```
OY      442 AGTGAATTCATTGCAATTGTATCTGGCTCAAATGGTCAACTAAGACGGCATC   501  
           ||| ||||| :||||| :|| |:: ||||| :: |||  
DB     88 ThrGlyAlaApIleValLeuGlueInLueGIunProVaLenHisSerSeArgVal    107  
                                           :: ||||| :|||  
OY     502 AAGAATGATCACTGCCACAACCCTCGAAATCCCGCACCTGGAACCTGTTGCATT   561  
           :: ||||| :|||  
DB    108 HisThrValMetLeuPProFroAlaSerGluThrPhetProGloglMetProCyStrPrval    127  
                                           :: ||||| :|||  
OY     562 GTTGTTATGGAAGAGCATGATAAGAC   588  
           |||:|||:  
DB    128 ThrgLyTrgDglYaepValAspAmnasp    136  
                                           |||:  
  
RESULT 10  
US-09-016-366A-17  
; Sequence 17, Application US/09016366A  
; Patent No. 5955431  
; GENERAL INFORMATION:  
; APPLICANT: Stevens, Richard L.  
; APPLICANT: Huang, Chifu  
; TITLE OF INVENTION: MAST CELL PROTEASE PEPTIDE  
; TITLE OF INVENTION: INHIBITORS  
; NUMBER OF SEQUENCES: 65  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Wolf, Greenfield & Sacks, P.C.  
; STREET: 600 Atlantic Avenue  
; CITY: Boston  
; STATE: MA  
; COUNTRY: U.S.A.  
; ZIP: 02210-2211  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: DOS  
SOFTWARE: FastSeq for Windows Version 2.0  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/016,366A  
FILING DATE: January 30, 1998  
CLASSIFICATION: 530  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 60/037,090  
FILING DATE: 05-FEB-1997  
ATTORNEY/AGENT INFORMATION:  
NAME: Plumer, Elizabeth R.  
REGISTRATION NUMBER: 36,637  
REFERENCE/DOCKET NUMBER: B0801/7093  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 617-720-3500  
TELEFAX: 617-720-2441  
TELEX:  
INFORMATION FOR SEQ ID NO: 17:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 275 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-09-016-366A-17  
  
Alignment Scores:  
Pred. No.:          0.000376         Length:       275  
Score:              119.00            Matches:        36  
Percent Similarity: 35.50%             Conservative:   24  
Best Local Similarity: 21.30%           Mismatches:    45  
Query Match:         4.43%               Indels:       64  
DB:                  2                 Gaps:          3
```

```
US-10-020-441-1 (1-1477) x US-09-016-366A-17 (1-275)  
OY      91 TGTAAGSTTCACTAGTCTCAACAGAGACAGTAGTCACTACAGACTGTGTTTT----- 144  
           ||| ||||| :||| :||| ||||| ||| |||||  
DB     59 CyseGlylgylserLeulleHisproGlnTrprValLeuthrAlahisScySleugLypPro   78  
           ---TGCTCACCATTTGCCAGTGATTTGGSGTAAAGAGATGACTGAAMACATTTGTTCTCA    201  
           145 --TGCTCACCATTTGCCAGTGATTTGGSGTAAAGAGATGACTGAAMACATTTGTTCTCA    201
```

Db 79 AspVallyAspLeuAlaThrLeuArgValGlnLeuArgGlnInHisLeuTyArgIn 98  
QY 202 TGTATCGATTGATTTCACAGACAGTGTGGTGCGTGCGTGTGTGTGTATGC 261  
Db 99 AspGln----- 100  
QY 262 AGTGAGTTTGACATGATTTTAATGCTATAPACTTGCTTATTTTCAGGTTTCATTTCTC 321  
Db 101 -----LeuLeuProValSerArgIleIleValHisProGlnPheTyArg 114  
QY 322 ACATGAGAGATGGCGACACACAGGACATCCATCACCAACCGTCTGGAGTTAAGTGCGCA 381  
Db 115 IleIleGln----- 117  
QY 382 CCAGATATCATGCCCTCTTGTATGTGCGACAGACAGAGAGACCAATGCGACAGACATC 441  
Db 117 ----- 117  
QY 442 AGTGATTCGATATTGCAATTGTAATGCTGCTCAATGTCATTAACAGATGGAATC 501  
Db 118 ThrGlyAlaAspIleAlaLeuLeuGlnLeuGlnProValAsnIleSerSerArgVal 137  
QY 502 AGATGATCAGTCTGCCACAGCCATCGATATCCGCCACCTGGAACTGGTGTTCATT 561  
Db 138 HistHrValMetLeuProAlaSerGlnThrPheProGlyMetProCysTrpVal 157  
QY 562 GTTGTTATGGAAGGATGATTAACGAC 588  
Db 158 ThrGlyTrpGlyAspValAspAsnAsp 166

## RESULT 11

US-08-978-404B-12

Sequence 12, Application US/08978404B

Patent No. 5968782

GENERAL INFORMATION:

APPLICANT: Stevens, Richard L.

TITLE OF INVENTION: MAST CELL PROTEASE THAT CLEAVES

TITLE OF INVENTION: FIBRINOGEN

NUMBER OF SEQUENCES: 74

CORRESPONDENCE ADDRESS:

ADDRESSEE: Wolf, Greenfield &amp; Sacks, P.C.

STREET: 600 Atlantic Avenue

CITY: Boston

STATE: MA

COUNTRY: U.S.A.

ZIP: 02210-2211

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette

COMPUTER: IBM Compatible

OPERATING SYSTEM: DOS

SOFTWARE: Pasteo for Windows Version 2.0

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/978,404B

FILING DATE: 25-NOV-97

CLASSIFICATION: 435

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 60/032,354

FILING DATE: 04-DEC-1996

ATTORNEY/AGENT INFORMATION:

NAME: Plumer, Elizabeth R.

REGISTRATION NUMBER: 36,637

REFERENCE/DOCKET NUMBER: B0801/7090

TELECOMMUNICATION INFORMATION:

TELEPHONE: 617-720-3500

TELEFAX: 617-720-2441

TELEX:

INFORMATION FOR SEQ ID NO: 12:

SEQUENCE CHARACTERISTICS:

LENGTH: 275 amino acids

TYPE: amino acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: No. 5968782e  
US-08-978-404B-12

## Alignment Scores:

Pred. No.:	0.000376	Length:	275
Score:	119.00	Matches:	36
Percent Similarity:	35.50%	Conservative:	24
Best Local Similarity:	21.30%	Mismatches:	45
Query Match:	4.43%	Indels:	64
DB:	2	Gaps:	3

US-10-020-441-1 (1-1477) x US-08-978-404B-12 (1-275)

QY 91 TGTACAGTTGACATGATTTTAATGCTATAPACTTGCTTATTTTCAGGTTTCATTTCTC 144  
Db 59 CysGlyGlySerLeuIleHisProGlnTrpValLeuThrAlaHisCysLeuIlePro 78  
QY 145 ---TGTACACATTTGCCAGTGTGTGGTAAAGATGCACTGAAACATTTCTCTCA 201  
Db 79 AspVallyAspLeuAlaThrLeuArgValGlnLeuArgGlnInHisLeuTyArgIn 98  
QY 202 TGTATCGATTGATTTCACAGACAGTGTGGTGCGTGCGTGTGTGTGTATGC 261  
Db 99 AspGln----- 100  
QY 262 AGTGAGTTTGACATGATTTTAATGCTATAPACTTGCTTATTTTCAGGTTTCATTTCTC 321  
Db 101 -----LeuLeuProValSerArgIleIleValHisProGlnPheTyArg 114  
QY 322 ACATGAGAGATGGCGACACACAGGACATCCATCACCAACCGTCTGGAGTTAAGTGCGCA 381  
Db 115 IleIleGln----- 117  
QY 382 CCAGATATCATGCCCTCTTGTATGTGCGACAGACAGAGAGACCAATGCGACAGACATC 441  
Db 117 ----- 117  
QY 442 AGTGATTCGATATTGCAATTGTAATGCTGCTCAATGTCATTAACAGATGGAATC 501  
Db 118 ThrGlyAlaAspIleAlaLeuLeuGlnLeuGlnProValAsnIleSerSerArgVal 137  
QY 502 AGATGATCAGTCTGCCACAGCCATCGATATCCGCCACCTGGAACTGGTGTTCATT 561  
Db 138 HistHrValMetLeuProAlaSerGlnThrPheProGlyMetProCysTrpVal 157  
QY 562 GTTGTTATGGAAGGATGATTAACGAC 588  
Db 158 ThrGlyTrpGlyAspValAspAsnAsp 166

## RESULT 12

US-09-079-970A-6

Sequence 6, Application US/09079970A

Patent No. 6274366

GENERAL INFORMATION:

APPLICANT: Maffitt, Mark A.

APPLICANT: Miles, Andrew L.

TITLE OF INVENTION: Enzymatically-Active Recombinant Human

TITLE OF INVENTION: Beta-Trypsin and Method of Making Same

NUMBER OF SEQUENCES: 6

CORRESPONDENCE ADDRESS:

ADDRESSEE: Intellectual Property Department

STREET: 8000 Excelsior Drive, Suite 401

CITY: Madison

STATE: WISCONSIN

COUNTRY: U.S.A.

ZIP: 53717-1914

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

```

1 APPLICATION NUMBER: US/09/079,970A
2 FILING DATE:
3
4 CLASSIFICATION: 435
5 ATTORNEY/AGENT INFORMATION:
6 NAME: Leone, Joseph T.
7 REGISTRATION NUMBER: 37,170
8 REFERENCE/DOCKET NUMBER: 34506.073
9 TELECOMMUNICATION INFORMATION:
10 TELEPHONE: (608) 831-2100
11 TELEFAX: (608) 831-2106
12 INFORMATION FOR SEQ ID NO: 6:
13 SEQUENCE CHARACTERISTICS:
14 LENGTH: 245 amino acids
15 TYPE: amino acid
16 TOPOLOGY: linear
17 MOLECULE TYPE: protein
18 US-09-079-970A-6
19
20 Alignment Scores:
21 Pred. No.: 0.00582 Length: 245
22 Score: 117.00 Matches: 37
23 Percent Similarity: 34.91% Conservative: 22
24 Best Local Similarity: 21.89% Mismatches: 46
25 Query Match: 4.36% Indels: 64
26 DB: 4 Gaps: 3
27
28 US-10-020-441-1 (1-1477) x US-09-079-970A-6 (1-245)
29
30 QY 91 TGTACAGSTTCAGTACTGTCTCAACAGACAGACAGTCTCAGCTGGTCAATGTTGTT----- 144
31 ||| |||||::: ||: ||||| ||||| |||||
32 Db 29 Cysgslgylserleuilllehlisproglintrpvalleuthralalahlscysvalglypro 48
33
34 QY 145 ---TGCACCAATTCGCCAGTATTCGGGTAAAGATCGATCGATGAAACATTTGTTGCTAA 201
35 ||| |||||::: ||: ||||| ||||| |||||
36 Db 49 AspvallysaspvleualalaleuhrygalgleuhrygluglnhlslenuYrYTYgln 68
37
38 QY 202 TGTATTCATGATTTCACACACACAGCTGTTCCGTGCGTGGTGTGTGTGTGTGTGTGTC 261
39 69 Aspqln----- 70
40
41 QY 262 AGTGGATTGCATAGATTTTATGATCTATTAATCTTACTTGGCTTATTTTCAGTTTCATTTCTC 321
42 ||| |||||::: ||: ||||| ||||| |||||
43 Db 71 -----LeuileuprovalserargllelevalhlsproglnpheYr 84
44
45 QY 322 ACACAGAGAAATGGCGAACCAAGGATCCATCCACCAACCGTGTGAGTAAGTGGCA 381
46 ||| ||: ||
47 Db 85 Thrallagln----- 87
48
49 QY 382 CCAGATATCATGCCCTCTTGTATGTGCGACAGACAGAGACCAATCGCACAGACATC 441
50 87 ----- 87
51
52 QY 442 AGTGATTCGATATTGCAATTGTATGCTGCTGCATAATGTCACATTCACAGAGGGAATC 501
53 ||| |||||::: ||: ||||| ||||| |||||
54 Db 88 lleglyalaspillealaleuileuglueudlnuprovallysvalserSerhlseval 107
55
56 QY 502 AGAGATGACAGTCTGCCACACGCCATCGGATATCCCGCACCTGGAGCATGGTGTTTCATT 561
57 ::::: |||||::: ||: ||||| ||||| |||||
58 Db 108 HlsthtrvalthlrleuprovalsercdlnhrphetprogllymetProcystrpval 127
59
60 QY 562 GTTGTTATGGAAGGATGATTAACGC 588
61 ||||| ||||| ||||| |||||
62 Db 128 ThrglytrpGlyaspvalaspAhnasp 136
63
64 RESULT 13
65 US-09-079-970A-5
66 Sequence 5, Application US/09079970A
67 Patent No. 6274366
68 GENERAL INFORMATION:
69 APPLICANT: Maffitt, Mark A.
70 APPLICANT: Niles, Andrew L.
71 APPLICANT: Haak-Frendescho, Mary
72 TITLE OF INVENTION: Enzymatically-Active Recombinant Human

```

```

1 TITLE OF INVENTION: Beta-Trypsinase and Method of Making Same
2 NUMBER OF SEQUENCES: 6
3 CORRESPONDENCE ADDRESSES:
4 ADDRESSEE: Intellectual Property Department
5 STREET: 8000 Excelsior Drive, Suite 401
6 CITY: Madison
7 STATE: WISCONSIN
8 COUNTRY: U.S.A.
9 ZIP: 53717-1914
10 COMPUTER READABLE FORM:
11 MEDIUM TYPE: Floppy disk
12 COMPUTER: IBM PC compatible
13 OPERATING SYSTEM: PC-DOS/MS-DOS
14 SOFTWARE: PatentIn Release #1.0, Version #1.30
15 CURRENT APPLICATION DATA:
16 APPLICATION NUMBER: US/09/079,970A
17 FILING DATE:
18 CLASSIFICATION: 435
19 ATTORNEY/AGENT INFORMATION:
20 NAME: Leone, Joseph T.
21 REGISTRATION NUMBER: 37,170
22 REFERENCE/DOCKET NUMBER: 34506.073
23 TELECOMMUNICATION INFORMATION:
24 TELEPHONE: (608) 831-2100
25 TELEFAX: (608) 831-2106
26 INFORMATION FOR SEQ ID NO: 5:
27 SEQUENCE CHARACTERISTICS:
28 LENGTH: 249 amino acids
29 TYPE: amino acid
30 TOPOLOGY: linear
31 MOLECULE TYPE: protein
32 US-09-079-970A-5
33
34 Alignment Scores:
35 Pred. No.: 0.00586 Length: 249
36 Score: 117.00 Matches: 37
37 Percent Similarity: 34.91% Conservative: 22
38 Best Local Similarity: 21.89% Mismatches: 46
39 Query Match: 4.36% Indels: 64
40 DB: 4 Gaps: 3
41
42 US-10-020-441-1 (1-1477) x US-09-079-970A-5 (1-249)
43
44 QY 91 TGTCAGSSTGCTACTAGTGTCTCAACAGACAGTACTCAACGCTGGTCAATTGTGT----- 144
45   ||| ||||| ||||| :||| ||||| |||||
46 Db 33 CysGIyGLysLeuIleHIsProGlInTrValIleuThrIalaIaHisCysValGIyPro 52
47
48 QY 145 ---TGCTACCATTTGCCAGTATTTCCGGGTAAAGATCGACTGGAACAACATTTGGTCTCAA 201
49   ||| ||||| ||||| :||| ||||| |||||
50 Db 53 AsPVallyAsPLeuAlaIalaIeuaIrVaIGlIndeuArfGuIglnIHisuTyTYrGln 72
51
52 QY 202 TGTATGCATGATTGATTTCACAGACAGTGTGGTCCGTGCGTGCCTGTGTGGTGGTAGTC 261
53   ||| ||||| ||||| :||| ||||| |||||
54 Db 73 AsgIn----- 74
55
56 QY 262 AGTGGTTTCATAGATTTTAATGCTATAATACCTTGCTTAATTTCAGGTTCAATTTC 321
57   ||| ||||| ||||| :||| ||||| |||||
58 Db 75 -----LeuIleProValSerarGILetIleValIHIsProGlnPhetyr 88
59
60 QY 322 ACACTGAGGAATGGCGAACCAACAAGGATCATCACCAACCGTGTGGAGTTMAAGTGGCA 381
61   ||| ||||| ||||| :||| ||||| |||||
62 Db 89 ThrIaGln----- 91
63
64 QY 382 CGAGGATACATGACCCTCTTGTATGTGGGACAGACAGAGAGAACAATGGACAGACATC 441
65   ||| ||||| ||||| :||| ||||| |||||
66 Db 91 ----- 91
67
68 QY 442 AGTGAATTGATATTGTCATATTGTATGCTGCTCAAAATGGTCAATTACAGAGTGAATC 501
69   ||| ||||| ||||| :||| ||||| |||||
70 Db 92 IlegIlaAsPrlleaIaleuIleuGlnIleuGlnIubProVallyValserSenHIsval 111
71
72 QY 502 AGAGTGAATCAAGTCTGCCACAGACCAATGGGATATCCCCGCACTTGGAGTGTGTTTCATT 561
73   ||| ||||| ||||| :||| ||||| |||||

```







```

NUMBER OF SEQUENCES: 65
CORRESPONDENCE ADDRESS:
ADDRESSEE: Wolf, Greenfield & Sacks, P.C.
STREET: 600 Atlantic Avenue
CITY: Boston
STATE: MA
COUNTRY: U.S.A.
ZIP: 02210-2211
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FASTSEQ for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/016,366A
FILING DATE: January 30, 1998
CLASSIFICATION: 530
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 60/037,090
FILING DATE: 05-FEB-1997
ATTORNEY/AGENT INFORMATION:
NAME: Plumer, Elizabeth R.
REGISTRATION NUMBER: 36,637
REFERENCE/DOCKET NUMBER: B0801/7093
TELECOMMUNICATION INFORMATION:
TELEPHONE: 617-720-3500
TELEFAX: 617-720-2441
TELEX:
INFORMATION FOR SEQ ID NO: 21:
SEQUENCE CHARACTERISTICS:
LENGTH: 274 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
US-09-016-366A-21

Alignment Scores:
Pred. No.: 0.00611          Length: 274
Score: 117.00             Matches: 37
Percent Similarity: 34.91%   Conservative: 22
Best Local Similarity: 21.89%  Mismatches: 46
Query Match: 4.36%          Indels: 64
Gaps: 3

US-10-020-441-1 (1-1477) x US-09-016-366A-21 (1-274)

QY 91 TGTACAGSTTCACTACTCTCAACAGAGAGATGACTACAGCTGGTATTGTGTT----- 144
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db 58 CysgIyglYserleuIlleHisProGIntPrValIleuThrAlaIahiscysValGlyPro 77
QY 145 ---TGCTACCAATGGCCAGATTCGGGTAAAGATCGACTGAACACATTTGTCCTCA 201
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db 78 AspValIyAspLeuAlaIaleuArgValGlnLeuArgGlnGlnHisLeuYrYrGln 97
QY 202 TGTAAATCATGATTATTCACAGACAGTGTTCGTCGTGCGCTTTGTGTGTATGTC 261
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db 98 AspGln----- 99
QY 262 AGTGGGTTGCATAGATTATTTATGCTATATTAATGCTCTTATTTACAGTTTCATTTCTC 321
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db 100 -----LeuLeuProValIserArgIleIleValHisProGlnPheYr 113
QY 322 ACACGTAGAAATGGCGACCAACAAGGANTCCATCCAACCCGTGGAAGTTAAGTGGA 381
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db 114 ThrAlaGln----- 116
QY 382 CCAGATACATGCCCTCTTGTATATGCGGACGACAGAGAGACCAATCGACAGACACTC 441
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db 116 ----- 116
QY 442 AGTGAATTCGATTTGCAATTTGTAATGCTGGCTCAAAATGCTCACTTACAGAGTGAATC 501
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db 117 IleGlyAlaAspIleAlaIleuLeuGlnGlnIleProValIyValIserSerHisVal 136

```

```

QY 502 AGAGTATCACTTCGCCACACCATCGATATCCCGCCACTGGAAGTGTGTTTCA 561
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db 137 HisThrValThrIleuProAlaIserGlnThrPheProGlyIleuProCysIlePrVal 156
QY 562 GTTGTTATGGAAGGATGATTAACGAC 588
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db 157 ThrGlyTrpGlyAspValAspAsnAsp 165

RESULT 18
US-08-978-404B-16
Sequence 16, Application US/08978404B
Patent No. 5968782
GENERAL INFORMATION:
APPLICANT: Stevens, Richard L.
TITLE OF INVENTION: MAST CELL PROTEASE THAT CLEAVES
TITLE OF INVENTION: FIBRINOGEN
NUMBER OF SEQUENCES: 74
CORRESPONDENCE ADDRESS:
ADDRESSEE: Wolf, Greenfield & Sacks, P.C.
STREET: 600 Atlantic Avenue
CITY: Boston
STATE: MA
COUNTRY: U.S.A.
ZIP: 02210-2211
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FASTSEQ for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/978,404B
FILING DATE: 25-NOV-97
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 60/032,354
FILING DATE: 04-DEC-1996
ATTORNEY/AGENT INFORMATION:
NAME: Plumer, Elizabeth R.
REGISTRATION NUMBER: 36,637
REFERENCE/DOCKET NUMBER: B0801/7090
TELECOMMUNICATION INFORMATION:
TELEPHONE: 617-720-3500
TELEFAX: 617-720-2441
TELEX:
INFORMATION FOR SEQ ID NO: 16:
SEQUENCE CHARACTERISTICS:
LENGTH: 274 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: No. 5968782e
US-08-978-404B-16

Alignment Scores:
Pred. No.: 0.00611          Length: 274
Score: 117.00             Matches: 37
Percent Similarity: 34.91%   Conservative: 22
Best Local Similarity: 21.89%  Mismatches: 46
Query Match: 4.36%          Indels: 64
Gaps: 3

US-10-020-441-1 (1-1477) x US-08-978-404B-16 (1-274)

QY 91 TGTACAGSTTCACTACTCTCAACAGAGAGATGACTACAGCTGGTATTGTGTT----- 144
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db 58 CysgIyglYserleuIlleHisProGIntPrValIleuThrAlaIahiscysValGlyPro 77
QY 145 ---TGCTACCAATGGCCAGATTCGGGTAAAGATCGACTGAACACATTTGTCCTCA 201
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db 78 AspValIyAspLeuAlaIaleuArgValGlnLeuArgGlnGlnHisLeuYrYrGln 97
QY 202 TGTAAATCATGATTATTCACAGACAGTGTTCGTCGTGCGCTTTGTGTGTATGTC 261

```

Db 98 AspGln----- 99  
Qy 262 AGTGGTTTGCATAGATTATTAAGTATATATCTTGCTTATTTCAGGTTTCATTTCTC 321  
Db 100 -----LeuLeuProValSerArgIleIleValHisProGlnPheTyr 113  
Qy 322 ACACAGGAGMATGGCGACCAACAGGATCATCAACCAACCGTTCGAGTTAAGTGGA 381  
Db 114 ThrAlaGln----- 116  
Qy 382 CCAGATACATCCCTCTTGTATGTCGGACACAGAGAGAACCAATCGCACAGACATC 441  
Db 116 ----- 116  
Qy 442 AGTGATTCGATTCGATTCGATTCGATTCGATTCGATTCGATTCGATTCGATTC 501  
Db 117 IlegIyAlaAspIleLeuLeuGluLeuGluProValIleValSerHisVal 136  
Qy 502 AGAGTATCATGTCGCACACAGCCATCGATTCGCCGCACATGGAATGTTGTTTCAT 561  
Db 137 H1StrValThleuProAlaSerGluThrPheProGlyMetProCysTrpVal 156  
Qy 562 GTTGTTATGGAAGGATATACAC 588  
Db 157 ThrGlyTrpGlyAspValAlaAspAsnAsp 165

## RESULT 19

US-08-944-483-58

; Sequence 58, Application US/08944483

; Patent No. 6232456

; GENERAL INFORMATION:

; APPLICANT: COHEN, MAURICE

; APPLICANT: COLPITTS, TRACEY L.

; APPLICANT: FRIEDMAN, PAULA N.

; APPLICANT: GRANADOS, EDWARD N.

; APPLICANT: KLAS, MICHAEL R.

; APPLICANT: RUSSELL, JOHN C.

; APPLICANT: STEWART, KENT D.

; APPLICANT: STROUPE, STEVEN D.

; TITLE OF INVENTION: NOVEL SERINE PROTEASE REAGENTS

; TITLE OF INVENTION: AND METHODS USEFUL FOR DETECTING AND TREATING DISEASES

; NUMBER OF SEQUENCES: 76

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Abbott Laboratories

; STREET: 100 Abbott Park Road

; CITY: Abbott Park

; STATE: IL

; COUNTRY: USA

; ZIP: 60064-3500

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Diskette

; COMPUTER: IBM Compatible

; OPERATING SYSTEM: DOS

; SOFTWARE: FastSeq for Windows Version 2.0

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/944.483

; FILING DATE:

; CLASSIFICATION: 424

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER:

; FILING DATE:

; ATTORNEY/AGENT INFORMATION:

; NAME: Becker, Cheryl L.

; REGISTRATION NUMBER: 35,441

; REFERENCE/DOCKET NUMBER: 6183, US.01

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: 847/935-1729

; TELEFAX: 847/938-2623

; TELEX:

; INFORMATION FOR SEQ ID NO: 58:

; SEQUENCE CHARACTERISTICS:

LENGTH: 242 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: NO. 6232456e  
US-08-944-483-58

## Alignment Scores:

Pred. No.:	0.00654	Length:	242
Score:	116.50	Matches:	37
Percent Similarity:	35.80%	Conservative:	21
Best Local Similarity:	22.84%	Mismatches:	49
Query Match:	4.34%	Indels:	55
DB:	4	Gaps:	2

US-10-020-441-1 (1-1477) x US-08-944-483-58 (1-242)

Qy 91 TGTACAGSTTCACTAGTCTCAACAGAGCAGTACTACAGCTGTGCTGCTTGTGCTCA 150  
Db 30 CysGlyGlySerLeuIleAlaProAspTrpValIleThrAlaGlyHisCysIleSerArg 49  
Qy 151 CCATTGCCAGTATTCGGTATAGATGACATGGAACACATTCGTCATATGATGCA 210  
Db 50 Aspleu----- 51  
Qy 211 TTGATTCACAGACAGTGTTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTC 270  
Db 51 ----- 51  
Qy 271 GCATGATTTTAAATGCTATATACCTGCTTATTCAGTTTCATTCACACTGACAG 330  
Db 52 -----ThyTyrGlnValIleValLeuGlyGluTyrAsnLeuAlaValIle 65  
Qy 331 AATGCGACCAACAGGATTCATCACCAACCGTGTGAGTTAAGTGGACACAGATAC 390  
Db 66 GluGlyProGluGlnValIleProIleAsnSerGluGlnLeuPheValHisProLeuTrp 85  
Qy 391 ATGCCCTCTGTATGTCGCACAGACAGAGAGACCAATGTCACAGACACTGATTC 450  
Db 86 AsnArgSerCysValAlaIle-----GlyAsn 94  
Qy 451 GATATTCATTTGATATGCTGCTGCTCAATGTCATTCATTCAGATGATGATGATC 510  
Db 95 AspleuAlaLeuIleLysLeuSerArgSerAlaGlnLeuGlyAspAlaValGlnLeuAla 114  
Qy 511 AGTCCGCACAGCCATCGATATCCGCCACTGGAATCGGTGTTTCAATGTTGTTAT 570  
Db 115 SerLeuProProAlaGlyAspIleLeuProAsnIleThrProCysTyrIleThrGlyTrp 134  
Qy 571 GGAAG 576  
Db 135 GlyArg 136

## RESULT 20

US-08-944-483-57

; Sequence 57, Application US/08944483

; Patent No. 6232456

; GENERAL INFORMATION:

; APPLICANT: COHEN, MAURICE

; APPLICANT: COLPITTS, TRACEY L.

; APPLICANT: FRIEDMAN, PAULA N.

; APPLICANT: GRANADOS, EDWARD N.

; APPLICANT: KLAS, MICHAEL R.

; APPLICANT: RUSSELL, JOHN C.

; APPLICANT: STEWART, KENT D.

; APPLICANT: STROUPE, STEVEN D.

; TITLE OF INVENTION: NOVEL SERINE PROTEASE REAGENTS

; TITLE OF INVENTION: AND METHODS USEFUL FOR DETECTING AND TREATING DISEASES

; NUMBER OF SEQUENCES: 76

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Abbott Laboratories

; STREET: 100 Abbott Park Road





MEDIUM TYPE: Diskette  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: DOS  
SOFTWARE: FastSeq for Windows Version 2.0  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/016,366A  
FILING DATE: January 30, 1998  
CLASSIFICATION: 530  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 60/037,090  
FILING DATE: 05-FEB-1997  
ATTORNEY/AGENT INFORMATION:  
NAME: Plumer, Elizabeth R.  
REGISTRATION NUMBER: 36,637  
REFERENCE/DOCKET NUMBER: B0801/7093  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 617-720-3500  
TELEFAX: 617-720-2441  
TELEX:  
INFORMATION FOR SEQ ID NO: 15:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 276 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-09-016-366A-15

Alignment Scores:  
Pred. No.: 0.00263 Length: 276  
Score: 111.00 Matches: 41  
Percent Similarity: 34.86% Conservative: 20  
Best Local Similarity: 23.43% Mismatches: 38  
Query Match: 4.13% Indels: 76  
DB: 2 Gaps: 5

US-10-020-441-1 (1-1477) x US-09-016-366A-15 (1-276)

QY 91 TGTACAGSTTCACTAGTCTCAACAGAGAGAGTCACTGAGTGTGTT----- 144  
DB 60 CysGlyGlySerLeuIleHisProGlnTrpValIleuThrAlaHisCysValGlyPro 79  
QY 145 -----TGCTCACATTGGCAGTGTTCGGTAAAGATCGACTGAACACATT----- 192  
DB 80 HisIleLysSerPro---GlnLeuPheArgValGlnLeuArgGlnGlnIleuTrpLeuTyrTyr 98  
QY 193 -----GTTGCTCAATGTAATGATTGATTTCACACAGAGTGTTCGGTGTGCTGC 243  
DB 99 GlyAspGlnLeuLeuSerLeuAsnArgIleVal----- 109  
QY 244 GTTTGTTGTTGTATGACGTGGGTTGCATGATTATTAATGTCATATTAATTGTTCTTA 303  
DB 109 ----- 109  
QY 304 TTTCAGTTTCACTTCTCACACTGAGAGATGGCGCAACAGGCATTCATCACCAACCG 363  
DB 109 ----- 109  
QY 364 TCTGAGTTAAGGTGCGACACAGATTCATGCCCTTGTATGTGCGCAGCAGAGAGAA 423  
DB 110 -----ValHisProHisTyrTyrThrAla----- 117  
QY 424 CCAATGCACAGACACTGATCGATTCGATTCGAAATTGTAATGCTGGCTCAATGTC 483  
DB 118 -----GlnGlyGlyAlaAspValAlaLeuLeuGlnLeuGlnValProVal 132  
QY 484 AACTTACAGAGTGAATCAGATGATCATCTGCCACAGCATCGATATCCGCCACCT 543  
DB 133 AsnValSerThrHisIleHisProIleSerLeuProAlaSerGlnThrPheProPro 152  
QY 544 GGAATGCTGTTTCACTGTTGTTGTTATGGAAGGATGATPACAGAC 588  
DB 153 GlyThrSerCysTrpValThrGlyTyrGlyAspIleAspAsnAsp 167

## RESULT 24

US-08-978-404B-21

Sequence 21, Application US/08978404B

Patent No. 5968782

GENERAL INFORMATION:

APPLICANT: Stevens, Richard L.

TITLE OF INVENTION: MAST CELL PROTEASE THAT CLEAVES

TITLE OF INVENTION: FIBRINOGEN

NUMBER OF SEQUENCES: 74

CORRESPONDENCE ADDRESS:

ADDRESS: Wolf, Greenfield & Sacks, P.C.

STREET: 600 Atlantic Avenue

CITY: Boston

STATE: MA

COUNTRY: U.S.A.

ZIP: 02210-2211

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette

COMPUTER: IBM Compatible

OPERATING SYSTEM: DOS

SOFTWARE: FastSeq for Windows Version 2.0

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/978,404B

FILING DATE: 25-NOV-97

CLASSIFICATION: 435

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 60/032,354

FILING DATE: 04-DEC-1996

ATTORNEY/AGENT INFORMATION:

NAME: Plumer, Elizabeth R.

REGISTRATION NUMBER: 36,637

REFERENCE/DOCKET NUMBER: B0801/7090

TELECOMMUNICATION INFORMATION:

TELEPHONE: 617-720-3500

TELEFAX: 617-720-2441

TELEX:

INFORMATION FOR SEQ ID NO: 21:

SEQUENCE CHARACTERISTICS:

LENGTH: 276 amino acids

TYPE: amino acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: No. 5968782e

US-08-978-404B-21

US-10-020-441-1 (1-1477) x US-08-978-404B-21 (1-276)

QY 91 TGTACAGSTTCACTAGTCTCAACAGAGAGTCACTGAGTGTGTT----- 144  
DB 60 CysGlyGlySerLeuIleHisProGlnTrpValIleuThrAlaHisCysValGlyPro 79  
QY 145 -----TGCTCACATTGGCAGTGTTCGGTAAAGATCGACTGAACACATT----- 192  
DB 80 HisIleLysSerPro---GlnLeuPheArgValGlnLeuArgGlnGlnIleuTrpLeuTyrTyr 98  
QY 193 -----GTTGCTCAATGTAATGATTGATTTCACACAGAGTGTTCGGTGTGCTGC 243  
DB 99 GlyAspGlnLeuLeuSerLeuAsnArgIleVal----- 109  
QY 244 GTTTGTTGTTGTATGACGTGGGTTGCATGATTATTAATGTCATATTAATTGTTCTTA 303  
DB 109 ----- 109  
QY 304 TTTCAGTTTCACTTCTCACACTGAGAGATGGCGCAACAGGCATTCATCACCAACCG 363







Db 157 GluAap-----ArgleuAenTyThrValIleAenSerThHisPheAenLeuThr 173  
 Qy 852 TCACGTCACCTCAGCTCA----- 835  
 Db 174 GluSerAenIleThSerIleGlnIleTyThrLeuAenThrThCysIleGluArgleuArg 193  
 Qy 834 -----CCAAACAATGACCGAGATGAAA 811  
 Db 194 AenTyThrLeuGluSerValTyThrThrThrValPro-GlnAenIleThThSerGI 213  
 Qy 810 TCACGTCGTCAGTCAAGTCGCAAGTCATCCAAACAGTTGTTATATCAAGCTCT 751  
 Db 213 nHisIaIeThrThrMetHisThrIleProAenThrIleThrIle----- 229  
 Qy 750 GACCAATACCAACCGGATCAGTAGACAGACAGACACAAGATACACAGACA 691  
 Db 230 -GlnAenThrThrGlnSerHisThrValGlnThrProSerPheAenAap-----ThrHi 247  
 Qy 690 CAGAAATACTAGTACATGCTGACTGACGACTGAGTGCCTTATATCAACAACACT 631  
 Db 247 sAsnValThrIyShIsthrLeuAenIleSerTyThrValLeuSer-GlnIySerThraAenA 267  
 Qy 630 CACTTCT 623  
 Db 267 hThrSer 269

RESULT 28

US-08-926-922-16  
 Sequence 16, Application US/08926922  
 Patent No. 5825751

GENERAL INFORMATION:  
 APPLICANT: Spaete, Richard  
 APPLICANT: Cha, Tai-An  
 TITLE OF INVENTION: NOVEL HUMAN CYTOMEGALOVIRUS  
 NUMBER OF SEQUENCES: 27  
 CORRESPONDENCE ADDRESS:  
 ADDRESSEE: Luann Cseerr Attorney at Law  
 STREET: 750 Arimo Avenue  
 CITY: Oakland  
 STATE: CA  
 COUNTRY: USA

ZIP: 94610  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: Floppy disk  
 COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: Patentin Release #1.0, Version #1.25  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/08/926,922  
 FILING DATE: September 10, 1997  
 CLASSIFICATION: 536  
 ATTORNEY/AGENT INFORMATION:  
 NAME: Cseerr, Luann  
 REGISTRATION NUMBER: 31,822  
 REFERENCE/DOCKET NUMBER: AVIR 11A  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: 510-834-1448  
 TELEFAX: 510-839-7810

INFORMATION FOR SEQ ID NO: 16:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 306 amino acids  
 TYPE: amino acid  
 TOPOLOGY: linear  
 MOLECULE TYPE: protein  
 IMMEDIATE SOURCE:  
 CLONE: tol.10  
 FEATURE:  
 NAME/KEY: ProteIn  
 LOCATION: 1..306  
 OTHER INFORMATION: /label= U142

US-08-926-922-16  
 Alignment Scores:

Pred. No.: 0.00728 Length: 306  
 Score: 107.00 Matches: 51  
 Percent Similarity: 38.73% Conservative: 28  
 Best Local Similarity: 25.00% Mismatches: 81  
 Query Match: 4.01% Indels: 44  
 DB: 2 Gaps: 6

US-10-020-441-1 (1-1477) x US-08-926-922-16 (1-306)

Qy 1152 CAAACAAACACCAAGTGAATTCGAAATATCAATGCGACAATATCCACCTTAC 1093  
 Db 84 GlnAenIySerTy-----AsnIySgInProleuGlyAerThrGlnThrLeuTy 100  
 Qy 1092 ACTACCCACACAAATAGTCCACATACATCAATGCACTATATTCAGATCAAGT 1033  
 Db 101 AsnIleAerSerGlnAenIleHisIleArgValSerGlnTyPheHisThrArgTyrIleIys 120  
 Qy 1032 CACTTCATCTCACCACCTCTCCAAACACATATCAACAAATGACCCCTACTTCATG 973  
 Db 121 SerleuGlnGluAenHisThrCysAerPleuThrAenSerThrProThr----- 136  
 Qy 972 TACACAAACACTTCATCGTTGATAGATATTCATAGATCTTCTCTCCAAAGTG 913  
 Db 137 TyrThrTyGlnValAenValAenAenThraenTyThrLeuThrThrSerSerIyTyr 156  
 Qy 912 TCTCATCATCCCATCATCATCATTCATCCCATCTCAATCTCAATCACTCAGTCA 853  
 Db 157 GluAap-----ArgleuAenTyThrValIleAenSerThHisPheAenLeuThr 173  
 Qy 852 TCACGTCACCTCAGCTCA----- 835  
 Db 174 GluSerAenIleThSerIleGlnIleTyThrLeuAenThrThCysIleGluArgleuArg 193  
 Qy 834 -----CCAAACAATGACCGAGATGAAA 811  
 Db 194 AenTyThrLeuGluSerValTyThrThrThrValPro-GlnAenIleThThSerGI 213  
 Qy 810 TCACGTCGTCAGTCAAGTCGCAAGTCATCCAAACAGTTGTTATATCAAGCTCT 751  
 Db 213 nHisIaIeThrThrMetHisThrIleProAenThrIleThrIle----- 229  
 Qy 750 GACCAATACCAACCGGATCAGTAGACAGACAGACACAAGATACACAGACA 691  
 Db 230 -GlnAenThrThrGlnSerHisThrValGlnThrProSerPheAenAap-----ThrHi 247  
 Qy 690 CAGAAATACTAGTACATGCTGACTGACGACTGAGTGCCTTATATCAACAACACT 631  
 Db 247 sAsnValThrIyShIsthrLeuAenIleSerTyThrValLeuSer-GlnIySerThraAenA 267  
 Qy 630 CACTTCT 623  
 Db 267 hThrSer 269

RESULT 29

US-09-253-682-16  
 Sequence 16, Application US/09253682  
 Patent No. 6040170

GENERAL INFORMATION:  
 APPLICANT: Spaete, Richard  
 APPLICANT: Cha, Tai-An  
 TITLE OF INVENTION: NOVEL HUMAN CYTOMEGALOVIRUS  
 NUMBER OF SEQUENCES: 27  
 CORRESPONDENCE ADDRESS:  
 ADDRESSEE: Luann Cseerr Attorney at Law  
 STREET: 750 Arimo Avenue  
 CITY: Oakland  
 STATE: CA  
 COUNTRY: USA

ZIP: 94610  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: Floppy disk  
 COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: PC-DOS/MS-DOS

```

1      SOFTWARE: Patent In Release #1.0, Version #1.25
2
3      CURRENT APPLICATION DATA:
4      APPLICATION NUMBER: US/09/253,682
5
6      FILING DATE:
7      CLASSIFICATION:
8      PRIOR APPLICATION DATA:
9      APPLICATION NUMBER: 08/926,922
10     FILING DATE: September 10, 1997
11     ATTORNEY/AGENT INFORMATION:
12     NAME: Cseer, Luann
13     REGISTRATION NUMBER: 31,822
14     REFERENCE/DOCKET NUMBER: AVIR 11A
15     TELECOMMUNICATION INFORMATION:
16     TELEPHONE: 510-834-1448
17     TELEFAX: 510-833-7810
18     INFORMATION FOR SEQ ID NO: 16:
19     SEQUENCE CHARACTERISTICS:
20     LENGTH: 306 amino acids
21     TYPE: amino acid
22     TOPOLOGY: linear
23     MOLECULE TYPE: protein
24     IMMEDIATE SOURCE:
25     CLONE: tol.10
26
27     FEATURE:
28     NAME/KEY: Protein
29     LOCATION: 1..306
30     OTHER INFORMATION: /label= U1142
31
32     US-09-253-682-16

```

Alignment Scores:	
Pred. No.:	0.00728
Score:	107.00
Percent Similarity:	38.73%
Best Local Similarity:	25.00%
Query Match:	4.01%
DB:	3
Gaps:	6
length:	306
Matches:	51
Conservative:	28
Mismatches:	81
Indels:	44
Gaps:	6

US-10-020-441-1 (1-1477) X US-09-253-682-16 (1-306)

[illegible]

Db 230 -GlnAsnThrThrGlnSerHisThrValGlnTyrProSerPheAsnAsp-----ThrHl 247

Qy 690 CAGATAATCTAGCATGTCAGATGACGTACGACTGATCATTCGTTATTTACCAACACT 631

Db 247 sAsnValThrIryHisThrLeuAnIleSerTyrValLeuSer-GlnLysThrAsnAsnT 267

Qy 630 CACTTTCT 623

Db 267 hrrmrSer 269

**RESULT 30**

US-09-527-657-16

: Patent NO. 6291236  
; Sequence 16, Application US/0932163/

GENERAL INFORMATION

APPLICANT: Spae

Cha, Tai-Am

TITLE OF INVENTION: NOVEL HUMAN CYTOMEGALOVIRUS  
NUMBER OF SENTENCES: 37

NUMBER OF SEQUENCES: 2/  
CORRESPONDENCE ADDRESS:

ADDRESSEE: Luann C

STREET: 750 Arimo Avenue

CITY: Oakland

STATE: CA  
COUNTRY:

COUNTRY: USA  
ZIP: 94610

COMPUTER READABLE

MEDIUM TYPE: FLOPPY

COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS

OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/0

FILING DATE: 17-Mar-2000

CLASSIFICATION: <U>  
; PRIOR APPLICATION DATA:

APPLICATION NUMBER: US/08/926,922

FILING DATE: September 1, 2011

ATTORNEY/AGENT INFORMATION

NAME: Cserr, Luann  
REGISTRATION NUMBER: 31 822

REFERENCE/DOCKET NUMBER:

TELECOMMUNICATION INFORMATION

TELEPHONE: 510-834-1

TELEFAX: 510-839-7810

SEQUENCE CHARACTERISTICS

SEQUENCE CHARACTERISTICS

LENGTH: 306 amino

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: protein  
IMMEDIATE SOURCE.

IMMEDIATE SOURCE:  
CLONE: F01.10

FEATURE: 004:40

US-09-527-657-16

Alignment Scores:

US-10-020-441-1 (1-1477) X US-09-527-657-16 (1-306)

QY 1152 CAAACAAAACACACAGTGATTTGAAAAATATCACTGCACAATATCCACCCCTTAC 1093  
|||||:::| | |||||  
Db 84 GlnAenlyserTyr-----AenlysglnProleuGlyAspThnGlnThrLeuTyr 100





[illegible][illegible]

ADDRESSEE: PETERS, VERNY, JONES & BIKSA  
STREET: 385 Sherman Avenue, Suite 6  
CITY: Palo Alto  
STATE: CA  
COUNTRY: USA  
ZIP: 94306-1840  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/928,361B  
FILING DATE: 12-SEP-1997  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 60/026,062  
FILING DATE: 13-SEP-1996  
ATTORNEY/AGENT INFORMATION:  
NAME: Verny, Hana  
REGISTRATION NUMBER: 30,518  
REFERENCE/DOCKET NUMBER: 480,76-1 (HV)  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 650-324-1677  
TELEFAX: 650-324-1678  
INFORMATION FOR SEQ ID NO: 27:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 216 amino acids  
TYPE: amino acid  
STRANDEDNESS:  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-928-361B-27

[illegible]

Score:	100.00	Matches:	44
Percent Similarity:	35.18%	Conservative:	26
Best Local Similarity:	22.11%	Mismatches:	53
Query Match:	3.72%	Indels:	76
DB:	2	Gaps:	6

US-10-020-441-1 (1-1477) x US-08-978-404B-5 (1-274)

```

QY      4 TCACACCTGGTGAATAGTAGGTGGAACCTGTGTGCACACCGCACTGAATTCCTGCATC 63
      ||| |||
Db      39 SerIyStr-----ProTrpGlnValSerLeuArgPheIysPheSer 52
QY      64 GCATTTCTTAACGACAGAGAGAACATGTGTACAGSTGTACATGATCTCAAGAGAGACGTA 122
      : : : : :
Db      53 PheTrpMet-----HisPheCysGlyCylSerLeuIleHisProGlnTrpVal 68
      : : : : :
QY      124 CTCACAGCTGGTCAATGTGTT-----TGCTCACCACTTGCACGATGATCGGTA 171
      ||||| ||||| |||||
Db      69 LeuThrIaIaHisCysValGlyLeuHisIleIysSerPro---GluLeuPheArgVal 87
QY      172 AGAATGCACTGAACAAACATCTGTGTCTCAATGTAACTGATGATGATTTCAACACAGTGT 231
      : : : : :
Db      88 GlnLeuArgGlnGlnTyrLeuTy------ 95
QY      232 GCGTGTGCGTCCGTTGTGTGTGTGTGTGTGATGACAGTGGCTTGTCATGATTTAAATGTCTATA 291
      : : : : :
Db      96 -----TyrIaIaArgGlnLeuThrVal 103
QY      292 TACTTGTCTTATTTACAGGTTTCATTTCTGCACACTGAGAAATGGCGACCAAGGCATC 351
      : : : : :
Db      104 AsnArgThrValValHisProHisIleTyrTyrThrValGluSer 117
QY      352 CATGACCAACCGTGTGAGATTAAAGTGGACCAAGATATCATGCCCTGTGTATGTGGCA 411
      : : : : :
Db      117 ----- 117
QY      412 CGACAGAGAGAACCAATCGACAGACACTGATGGATTCGATATTTGCAATTTGTAATGCTG 471
      : : : : :
Db      118 -----GlyIaAspIleAlaLeuGluLeu 126
QY      472 GCTCAATGTGCACTTACAGAGTGAATGAGATGATGATGATGCTGCACAGCCATCGAT 531
      ||||| : : : : :
Db      127 GluIleProValIaHisValSerThrHisIleHisProIleSerLeuProProIaIaSerGlu 146
QY      532 ATCCCGCACCTGTGAAGTGTGTGTTTCATTTGTTGTTATGGAAGGATGAACGAC 588
      ||| : : : : :
Db      147 ThrPheProSerGlyThrIleSerCysTrpValThrGlyTrpGlyAspIleAspSerAsp 165

RESULT 38
US-08-978-404B-3
: Sequence 3, Application US/08978404B
: Patent No. 5968782
: GENERAL INFORMATION:
: APPLICANT: Stevens, Richard L.
: TITLE OF INVENTION: WAST CELL PROTEASE THAT CLEAVES
: TITLE OF INVENTION: FIBRINOGEN
: NUMBER OF SEQUENCES: 74
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Wolf, Greenfield & Sacks, P.C.
: STREET: 600 Atlantic Avenue
: CITY: Boston
: STATE: MA
: COUNTRY: U.S.A.
: ZIP: 02210-2211
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Diskette
: COMPUTER: IBM Compatible
: OPERATING SYSTEM: DOS
: SOFTWARE: FASTSEQ for Windows Version 2.0
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/978,404B
: FILING DATE: 25-NOV-97
: CLASSIFICATION: 435

```

```

; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 60/032,354
; FILING DATE: 04-DEC-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Plumer, Elizabeth R.
; REGISTRATION NUMBER: 36,637
; REFERENCE/DOCKET INFORMATION:
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617-720-3500
; TELEFAX: 617-720-2441
; TELEX:
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 273 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: No. 5968782e
; US-08-978-404B-3

```

```

Alignment Scores:
Pred. No.: 0.0485 Length: 273
Score: 99.00 Matches: 37
Percent Similarity: 34.33% Conservative: 28
Best Local Similarity: 19.58% Mismatches: 56
Query Match: 3.69% Indels: 68
DB: 2 Gaps: 4

```

US-10-020-441-1 (1-1477) x US-08-978-404B-3 (1-273)

```

QY 40 CACCGCACTGAATCCCGTTCATCGATTCTTAACACAGAGAGAA----- 87
   |||
   |||
Db 36 HSGLYAAsnlystrprtrprglInValserLeuArgAlaasnAspThrYrtrpMeth: 55
   |||
QY 88 ATGTGACAGAGTTCTAATCTCAAGAGAGAGTCTACAGCTGTGCTATGTTGCT 147
   |||
   |||
Db 56 PheCysGlyGlySerIleuLeuHisProGlnTrpValIleuThrAlaAlaHisCysValGly 75
   |||
QY 148 TCACCATTCGCA-----GTGATTCGGGTAAAGATCGACTGAACACATTTGTGCT 198
   |||
   |||
Db 76 ProAspValAlaAspProAsnlyserValArgValGlnLeuArgIleGlnIleuYrtrp 95
   |||
QY 199 CAATGTAATCGATTGATTTTCACAGACAGTGTTCGCTGCGTGTGCTTTGTGTGA 258
   |||
   |||
Db 96 HisAspHis----- 98
   |||
QY 259 TGCAGTGGGTTGCATAGATTTTAATGTCATATACTGTGCTTATTTACAGTTTCA 318
   |||
   |||
Db 99 -----LeuMetThrValSerGlnIleLeuThrHisProAspPhe 111
   |||
QY 319 CTCACACTGAGAGATGCGACCAACAAGCATCATCACCAACCGTCTGGAGTTAAGT 378
   |||
   |||
Db 112 TyrIleValGlnAsp----- 116
   |||
QY 379 GCACACAGATACATGCCCTCTTGTAATGTCGGACGACAGAGAGACCAATCGCACAGACA 438
   |||
   |||
Db 116 ----- 116
QY 439 CTCAGTGCATGATTTGCAATTTGTAATGCTGCTCAATGTCTCAACTTACAGAGTGA 498
   |||
   |||
Db 117 -----GlyAlaAspIleAlaLeuLeuLysLeuThrAsnProValAsnIleSerAspTyr 134
   |||
QY 499 ATCAGAGTATCATGCTGCCACAGCCATCGATATCCCGCACCTGGAACCTGGTGTTC 558
   |||
   |||
Db 135 ValHisProValProLeuProPheAlaSerGlnThrPheProSerGlyThrIleuCysTrp 154
   |||
QY 559 ATTGTTGTTTGAAGAGGATGATAAC 585
   |||
   |||
Db 155 ValThrGlyTyrGlyAsnIleAspAsn 163
   |||
   |||

```

RESULT 39  
US-09-386-653A-7  
; Sequence 7, Application US/09386653A

```

; Patent No. 6458564
; GENERAL INFORMATION:
; APPLICANT: Andrade-Gordon, Patricia
; APPLICANT: Darrow, Andrew
; APPLICANT: Qi, Jian-shen
; TITLE OF INVENTION: DNA encoding the novel human serine
; FILE OF INVENTION: procase 1
; FILE REFERENCE: ORT-1032
; CURRENT APPLICATION NUMBER: US/09/386,653A
; CURRENT FILING DATE: 1999-08-31
; NUMBER OF SEQ ID NOS: 11
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 7
; LENGTH: 290
; TYPE: PRT
; ORGANISM: Homo sapiens
; US-09-386-653A-7

```

```

Alignment Scores:
Pred. No.: 0.0498 Length: 290
Score: 99.00 Matches: 49
Percent Similarity: 36.90% Conservative: 20
Best Local Similarity: 26.20% Mismatches: 48
Query Match: 3.63% Indels: 70
DB: 4 Gaps: 9

```

US-10-020-441-1 (1-1477) x US-09-386-653A-7 (1-290)

```

QY 91 TGTACAGSTTCACTGCTGCTCAACGAGACGATCTACAGCTGCTGATGTTGCTCA 150
   |||
   |||
Db 60 CysGlyGlySerIleAlaGlnIleTrpValIleuThrAlaAlaHisCysPhe----- 77
   |||
QY 151 CCATTGCCAGTATTCGGGTAAAGATGACGACTGAACAACATTTGTTCTCAATGTAATGCA 210
   |||
   |||
Db 78 -----Arg 78
   |||
QY 211 TTGATTTACAGACAGCTGTTGCGTGCCTGTTGTGTTGTGTAATGACAGTGGCTTT 270
   |||
   |||
Db 79 AsnThrSerGlnThr----- 83
   |||
QY 271 GCATGATTTTAATGCTTAATATCTGCTTATTTTCAAGTTTCAATTTCTACACTGAGG 330
   |||
   |||
Db 84 -----SerLeuTyrGlnValIleuLeu----- 90
   |||
QY 331 AATGCCGACCAACAAGATCCATCATCAACCCGTCGAGTTAAGTGGACACAGATATAC 390
   |||
   |||
Db 91 ---GlyAlaArgGln-----LeuValGlnProGly----- 99
   |||
QY 391 ATGCCCTCTTGTATTCGGCAGCAGACAGAGAG-----CAATCCACAG----- 435
   |||
   |||
Db 100 ---ProHisAlaMetTyrAlaArgValArgIleValGlnSerAsnProLeuTyrGlnGly 118
   |||
QY 436 ACACTCAGTGGATTCGATTTGCAATTTGTAATGCTGCGCTCAATGTCTCAACTTACAGAGT 495
   |||
   |||
Db 119 ThrAlaSerSerAlaAspValAlaLeuValGlnLeuGlnAlaProValProPheThrAsn 138
   |||
QY 496 GGAATCAGATGATCATGCTGCCACAGCCATCGATATCCCGCACCTGGAACCTGATGTT 555
   |||
   |||
Db 139 TyrIleLeuProValCysLeuProAspProSerValIlePheGlnThrGlyMetAsnGly 158
   |||
QY 556 TTCATTTGTTGTTATGAGAG---GATGATTAACGACCGTATCCCTCACGTAAGAAATGCT 612
   |||
   |||
Db 159 TrpValThrGlyTyrGlySerProSerGlnIleAspLeuProGlnProArg----- 176
   |||
QY 613 GGAATTTGAAGAAAGTAGT 633
   |||
   |||
Db 177 ---IleLeuGlnLysIleuAla 182
   |||
   |||

```

RESULT 40  
US-09-386-653A-9  
; Sequence 9, Application US/09386653A  
; Patent No. 6458564  
; GENERAL INFORMATION:



APPLICANT: Andrade-Gordon, Patricia  
APPLICANT: Darrow, Andrew  
APPLICANT: Qi, Jian-shen  
TITLE OF INVENTION: DNA encoding the novel human serine  
TITLE OF INVENTION: Protease T  
FILE REFERENCE: ORT-1032  
CURRENT APPLICATION NUMBER: US/09/386,653A  
CURRENT FILING DATE: 1999-08-31  
NUMBER OF SEQ ID NOS: 11  
SOFTWARE: PatentIn Ver. 2.0  
SEQ ID NO: 9  
LENGTH: 315  
TYPE: PRT  
ORGANISM: Artificial Sequence  
FEATURE:  
OTHER INFORMATION: Description of Artificial Sequence: Fusion Protein  
OTHER INFORMATION: of Protease T in a zymogen activation construct  
US-09-386-653A-9

Alignment Scores:  
Pred. No.: 0.0516 Length: 315  
Score: 99.00 Matches: 49  
Percent Similarity: 36.90% Conservative: 20  
Best Local Similarity: 26.20% Mismatches: 48  
Query Match: 3.69% Indels: 70  
DB: 4 Gaps: 9

US-10-020-441-1 (1-1477) x US-09-386-653A-9 (1-315)

QY 91 TGTACAGTTCCTAGTCTCAACAGAGCAGTACTCAGCTGCTGCTGCTGCTCA 150  
Db 77 CysGlyGlySerLeuIleAlaGluIntrValIleThrAlaIleCysPhe----- 94  
QY 151 CCATTGCCAGTATTCGGGTAAAGATGACATGAAACATTTGCTCAATGTATCGA 210  
Db 95 -----Arg 95  
QY 211 TTGATTCACAGACAGTGTTCGTCGCGTGTGTTGTGTATGACAGTGGTTT 270  
Db 96 AsnThrSerGluThr----- 100  
QY 271 GCATAGATTTTAAATGCTATATATCTTGCTTATTTCAAGTTTCAATTTCTCACTGAG 330  
Db 101 -----SerLeuTyrglnValLeu----- 107  
QY 331 AATGCCAGCAACAGAGCATCATCACCACCGCTGAGTTAAGTGCACACAGATAC 390  
Db 108 --GlyAlaArgGln-----LeuValGlnProGly--- 116  
QY 391 ATGCCCTCTTGTATGTCGACAGACAGAGAGA-----CCAAATCGCAGAG--- 435  
Db 117 ---ProHsAlaMetTyrlAlaArgValArgInValGlnSerAsnProLeuTyrglnGly 135  
QY 436 ACACACTGAGTTCGATATTCGAAATTTGATGCTGCTCAATGATGTCAACTTACAGAGT 495  
Db 136 ThrAlaSerSerAlaAspValAlaLeuValGlnLeuGlnAlaProValProPheThrAsn 155  
QY 496 GGAATCAAGATGATCAGTCTGCCACAGCCATGCAATATCCGCCACCTGGAACCTGTGTT 555  
Db 156 TyrlLeuProValCysLeuProAspProSerValIlePheGluThrGlyMetAsnCys 175  
QY 556 TTCATTGTTGTTATGGAAG---GATGATACGACCGGTACCGTACGTAAGATGTT 612  
Db 176 TrpValThrGlyTrpGlySerProSerGlnGluAspLeuLeuProGluProArg----- 193  
QY 613 GGAATATTGAAGAAAGTGAAT 633  
Db 194 ---IleLeuGlnIleLeuAla 199

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.4 p5.4578  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - protein search, using frame\_plus\_n2p model

Run on: April 9, 2003, 13:26:44 ; Search time 61 Seconds  
(without alignments)  
4655.428 Million cell updates/sec

Title: US-10-020-441-1

Perfect score: 2686  
Sequence: 1 ggtcacactggtgacacg.....ctgtacgctcccaatttga 1477

Scoring table: BLOSUM62

Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 283224 seqs, 96134422 residues

Total number of hits satisfying chosen parameters: 566448

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Command line parameters:

-MODEL=frame+n2p model -DEV=xlp  
-Q=/cgnr2/1/USFPO/US10020441/runat 09042003 121212 1354/app query.fasta\_1.1671  
-DB=PIR\_73 -QFMT=fastcan -SUFFIX=rrp -MINMATCH=0.1 -LOOPEL=0 -LOOPEXT=0  
-UNITS=bits -START=1 -END=1 -MATRIX=biosum62 -TRANS=human40.cdi -LIST=45  
-DOCALLIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=40 -MODE=LOCAL  
-OUTFMT=pct -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000  
-USRB=US10020441 @CGN 1 1 80 @runat 09042003 121212 1354 -NCP=6 -ICPU=3  
-NO\_XLIPPY -NO\_MMAB -LARGEQUERY -NEG\_SCORES=0 -WAIT -LONGLOG -DEV TIMEOUT=120  
-MARK TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -Fgapop=6 -Fgapext=7  
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

1: PIR\_73:\*  
2: PIR2:\*  
3: PIR3:\*  
4: PIR4:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1091.5	40.6	264	2	A28942 pancreatic elastase
2	138.5	5.2	610	2	S35049 mucin JERS7 - huma
3	124.5	4.7	3570	2	T45025 mucin MUC5B, trach
4	124	4.6	237	2	S68702 trypsinase (EC 3.4.2
5	123	4.6	275	2	A35863 trypsinase (EC 3.4.2
6	123	4.6	275	2	C35863 trypsinase (EC 3.4.2
7	121.5	4.6	505	2	B46629 mucin 6, gastric (
8	121	4.5	281	2	T13596 trypsin homolog -
9	120	4.5	477	2	S53362 mucin 5AC (clone J
10	118	4.4	270	2	A29934 pancreatic elastase
11	117	4.4	270	2	S56160 mast cell trypsinase
12	117	4.4	275	2	B35863 mast cell trypsinase
13	115.5	4.3	1161	2	S57180 probable membrane
14	114.5	4.3	270	2	B29934 pancreatic elastase

C 15	113	4.2	279	2	S54157	extensin-like prot
C 16	113	4.2	351	1	KG2QHL	histidine-rich gly
C 17	112.5	4.2	273	2	T44657	protein GP80 (limp
C 18	111	4.1	230	2	I48685	mast cell proteina
C 19	111	4.1	276	2	A38654	mast cell proteina
C 20	110.5	4.1	367	2	G85352	hypothetical prote
C 21	109.5	4.1	274	2	A45754	trypsinase (EC 3.4.2
C 22	109	4.1	770	2	T22808	hypothetical prote
C 23	108.5	4.1	121	2	G82853	hypothetical prote
C 24	108.5	4.1	3119	2	T18414	protein g377 - mal
C 25	107.5	4.0	370	1	W2W139	E2 protein - human
C 26	107.5	4.0	1832	2	T31113	mucin-like glycopr
C 27	107	4.0	140	2	A54523	histidine-rich pro
C 28	107	4.0	246	2	A38678	mast cell proteina
C 29	107	4.0	786	2	T16509	hypothetical prote
C 30	106	4.0	6420	2	T30283	polypeptide synthas
C 31	105.5	4.0	203	2	B27059	hypothetical prote
C 32	104.5	3.9	330	2	S37595	mucin JUI10 - huma
C 33	104.5	3.9	1118	2	A48292	mucin, tracheobron
C 34	104.5	3.9	3020	2	A43932	pancreatic elastase
C 35	104	3.9	266	1	ELRT1	hypothetical prote
C 36	104	3.9	709	2	T28712	hypothetical prote
C 37	104	3.9	1630	2	A53577	acities siatoglyco
C 38	104	3.9	2225	2	T26063	hypothetical prote
C 39	103.5	3.9	560	2	S50392	hypothetical prote
C 40	103.5	3.9	651	2	C56653	membrane glycoprot
C 41	103.5	3.9	1180	2	S69205	stripe a/b protein
C 42	103.5	3.9	1817	2	D71606	hypothetical prote
C 43	103	3.9	660	2	UW067	chitinase (EC 3.2.
C 44	103	3.9	1169	2	S38181	floculation prote
C 45	102.5	3.8	764	2	H71607	hypothetical prote

ALIGNMENTS

RESULT 1  
A28942 pancreatic elastase (EC 3.4.21.36) precursor - fluke (Schistosoma mansoni)  
C:Species: Schistosoma mansoni  
C:Date: 30-Jun-1989 #sequence\_revision 30-Jun-1989 #text\_change 26-Aug-1999  
R:Accession: A28942  
R:Newport, G.R.; McKerrow, J.H.; Hedstrom, R.; Pettit, M.; McGarrigle, L.; Barr, P.J.;  
J. Biol. Chem. 263, 13179-13184, 1988  
A:Title: Cloning of the proteinase that facilitates infection by schistosome parasites.  
A:Reference number: A28942; MUID:88330818; PMID:33166457  
A:Accession: A28942  
A:Molecule type: mRNA  
A:Residues: 1-264 <NEW>  
A:Cross-references: GB:J03946; NID:g160947; PID:g160948  
C:Superfamily: trypsin; trypsin homology  
C:Keywords: hydrolase; serine proteinase  
F:27-256/Domain: trypsin homology <TR>

Alignment Scores:

Pred. No.: 1.79e-90 Length: 264  
Score: 1091.50 Matches: 239  
Percent Similarity: 48.58% Conservative: 0  
Best Local Similarity: 48.58% Mismatches: 3  
Query Match: 40.64% Indels: 250  
DB: Gaps: 2

US-10-020-441-1 (1-1477) x A28942 (1-264)

QY 1 GTGTCAACCTGCTGATTCGATGCGTCAACCTGTCACACCGCACTGAATTCGGCTTC 60  
DB 23 ValSerThrTrpLeuIleArgSerGlyGluProValGlnHisProValGluPheProPhe 42  
QY 61 ATTCGATTCCTTAACGACAGAGAACAATGTCACAGTTCACTAGTCTCAACGAGACA 120  
DB 43 IleAlaPheLeuThrThrGluArgThrMetCysThrGlySerLeuValSerThrArgGla 62  
QY 121 GTACTACAGCTGCTGATTCATTTGTTTGCTCAACATTGCACAGTTCGGGTGAAGATCGA 180

```

Db      63  ValLeuThrAlaGlyHisCysValCysSerProLeuProValIleArg----- 78
Qy      181  CTGAACAACATGTTGCTCATATGTAATGATTTTCACAGACAGTGTTCGCTGTGGC 240
Db      78  ----- 78
Qy      241  TGGCTTTGTTTGTGTATGACAGTGGGTTTGACATGATTTTAATGCTATATACCTTGTC 300
Db      78  ----- 78
Qy      301  TTATTTACAGTTTCATTTCTCACTAGAGAAATGGCGAACAGCATTCATCACCA 360
Db      79  -----ValSerPheLeuThrLeuArgAsnGlyAspGlnGlnGlyIleHisGln 95
Qy      361  CGCTCGAGAGTTAAAGGTGGCACAGGATACATGCCCTCTGATGCGGCACAGACAGAG 420
Db      96  ProSerGlyValAllyValAlaProGlyIlyrMetProSerCysMetSerAlaArgGlnArg 115
Qy      421  AGACCAATCGCACAGACACTCAGTGATTCGATATGCAATTTGTAATGCTGGCTCAATG 480
Db      116  ArgProIleAlaGlnThrLeuSerGlyPheAspIleAlaIleValMetLeuAlaGlnMet 135
Qy      481  GTCAACTTACAGAGTGAATCAGAGTCAAGTCTGCCACAGCCATCGGATATCCGGCCA 540
Db      136  ValAsnLeuGlnSerGlyIleArgValIleSerLeuProGlnProSerAspIleProPro 155
Qy      541  CCTGGAACAGTGTGTTTCATTTGTTGGTTATGGAAGGATATACAGCCGTGATCCGCTCA 600
Db      156  ProGlyThrGlyValAlaPheIleValGlyIlyrGlyArgAspAspAsnAspArgAspProSer 175
Qy      601  CGTAAAGATGTGGAATATATTAAGAAAGTGAAGTGTGTTGGTGAATAAACACATGACTCAG 660
Db      176  ArgLysAsnGlyGlyIleLeuLys----- 184
Qy      661  TCGTCAGTCAGATGTCAGTCAGTATTCCTGTGTCTGTCTGTATCTGTGTGTCTGTCTGTC 720
Db      184  ----- 184
Qy      721  TGTCTACCTGATCCGGTGTGTTGATTTGTCAGAGCCTTGATTAACAACCTGTTTGA 780
Db      184  ----- 184
Qy      781  TCACTTTTGACAGTTCACTAGACAGAGTGAATTTCCATCTCGGTCATTTGTTGGTAGGT 840
Db      184  ----- 184
Qy      841  GAGGTGACGTGATGTCAGGTGAGTTGAGTGAATTTGGATGGATGGAATGTGATGTGATG 900
Db      184  ----- 184
Qy      901  GGATGATTTGACACACTTGGAGAGAGAAAGATCATGAATAATCTATGCAACGATGAA 960
Db      184  ----- 184
Qy      961  GTGTGTGTGTACATGAAAGTGAAGGGGTCAATGTGTTGAGTATGTGTTGAGAGTGTG 1020
Db      184  ----- 184
Qy      1021  AGATGAGAGTGACTTTGATCGTGAATATATGTCACATGTGATTTGATGTGACATAATGT 1080
Db      184  ----- 184
Qy      1081  TGTGTGGGTAGTGAAGGTGAGTATTTGTGCCAGTTGATTTTGGAAATTCACTTGG 1140
Db      184  ----- 184
Qy      1141  TGTGTTGTTTGTTCCTGTTGTTGTTTGTCTTTCATCTGATCTGATCTGATCTGATCTG 1200
Db      184  ----- 184
Qy      1201  GATCTGTTGTGTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTT 1260
Db      185  -----GlyArgAlaThrIleMetGlnCysArg 193

```

```

Qy      1261  ACATGCGACCAATGGCAATCTCTATATGTGTGAAAGCAGGTGACAGAAATTTGACACTTACC 1320
Db      193  GHisAlaThrAsnGlyAsnProIleCysValIlyleAlaGlyGlnAsnPheGlyGlnLeuPro 213
Qy      1321  CGCTCAGGTGACAGTGTGGACCTCTCTCCATCTCTTCAAGGTCCAGTATCTCGGTGT 1380
Db      213  AlaProGlyAspSerGlyGlyProLeuLeuProSerLeuGlnGlyProValLeuGlyVal 233
Qy      1381  CGTATCACATGTTGTCACACTGCGCAACCTTCCGATATCATTTGTGAGTATGACAGTGT 1440
Db      233  ValSerHisGlyValThrLeuProAsnLeuProAspIleIleValGlnIlyrAlaSerVal 253
Qy      1441  GGTAGAAATGTTGATTTTGTACGCTCCAAATAT 1474
Db      253  ValArgMetLeuAspPheValArgSerAsnIle 264

RESULT 2
S35049
mucin JER57 - human (fragment)
C.Species: Homo sapiens (man)
C.Date: 10-Dec-1993 #sequence_revision 20-Apr-2000 #text_change 20-Apr-2000
C.Accession: S35049, S37594
R.DuToise, J.; Porchet, N.; Audie, J.P.; Guyonnet Duperrat, V.; Iaine, A.; van-Seuningen
Biochem. J. 293, 329-337, 1993
A.Title: Degenerate 87-base-pair tandem repeats create hydrophilic/hydrophobic alternat
A.Reference number: S35047; MUID:93343858; PMID:7916618
A.Accession: S35049
A.Status: preliminary
A.Molecule type: mRNA
A.Residues: 1-610 <DUF>
R.Aubert, J.
submitted to the EMBL Data Library, September 1993
A.Reference number: S37593
A.Accession: S37594
A.Molecule type: mRNA
A.Residues: 1-20, 'w', 22-610 <AUB>
A.Cross-references: EMBL:X74955

Alignment Scores:
Pred. No.: 0.000333 Length: 610
Score: 138.50 Matches: 67
Percent Similarity: 38.74% Conservative: 31
Best Local Similarity: 26.48% Mismatches: 78
Query Match: 5.20% Indels: 77
DB: 2 Gaps: 12

US-10-020-441-1 (1-1477) x S35049 (1-610)
Qy      1369  TGAACCTTGAAGGATGGAGGAGGATGCACACCTGTCACTGGAGCGGATTAAGTCC 1310
Db      11  TrpCysProIlePdlYSerThrAlaArgSerThrAsnAlaIleTrp----- 25
Qy      1309  GAAATTTGACCTGCTTTTCAACATATAGATTTGSCATTTGGTGGATGTGGCATTCAT 1250
Db      26  -----ThrLeuAlaTrpSerAlaGlyThrValSerArgTrpGly 38
Qy      1249  TATAGTGC-----TGCACCTG-----TGTGCACAAACAAACACACACACAGTA 1202
Db      39  SerSerArgCysAlaSerThrMetLysSerValCysSerAlaAlaIleThrProLeuPro 58
Qy      1201  CAGTCAACAAACGTACAGTACAGATGACAAAGAACAAACAAACACACGSA----- 1154
Db      59  GlnHisProGlyHisGlnLeuIlyrAlaMetProSerSerThrProGlyThrThrIlePro 78
Qy      1153  ---ACAAAACAAACACACACAGTGAATTTGCAAAATATCACTGGACATATTCACCTT 1097
Db      79  LeuThrGlnLeuThrThrThrAspHisIlyrAspCysIleHisIleTrpIleHisGlyHisPro 98
Qy      1096  TTACACTACCCACACACATATGATGACATATACATACACATGAGTGAATATTCAGCATC 1037
Db      99  ValLeuHisProGlyAsnSerSerPro-----Ser-GlnSe 110

```

QY 1036 AAGTCACTCTCATCTCAACCACTCTCAAAACATATCAACATGACCCCTACTT 977  
 Db 110 rAlaAspInProGlyHisHis----- 117  
 QY 976 CATGACACACACACTTCATCGTTTGATGATATTTTCAAGAGTCTTCTCTCCAA 917  
 Db 118 -----ThrHisSerHisGlnPheGlnSerHisPhe-----LeuLeuGlnSerHis 132  
 QY 916 GTGGTCAATCATCCCATCATCATCA-----CATTC----- 885  
 Db 132 s-AspCysAsnHisProSerSerHisAlaAspAsnSerHisGlnLeuHisSerTyrGlnV 152  
 QY 884 -----ATCCCATCCCAAT-----CTACCTCAACTCACTCAAC 854  
 Db 152 aLThrProIleProSerSerHisThrLeuGlyThrThrGlyThrLeuProGluGlnThrHis 172  
 QY 853 ATCAGCTCACTCACTCAACACACATGACCGAGATGGAATCATCTGTCTACTGAAC 794  
 Db 172 IsProValAlaThr-----MetSerThrIleHisProSerSerThrP 186  
 QY 793 TGTCAACAGTCACTCAACACAGTTGTTATTATCAAGGCTGTGACCAATACAAACCG 734  
 Db 186 roGluThrThrHisThrSerThrValLeuThrThrLysAlaThrThrAlaThrHis 206  
 QY 733 GATCAGTAGACAGACAGACACAAACAGATACA 699  
 Db 206 erSer-----ThSerThrProSerSerThr 214

## RESULT 3

T45025  
 mucin MUC5B, tracheobronchial [imported] - human (fragment)  
 C/Species: Homo sapiens (man)  
 C/Date: 21-Jan-2000 #sequence\_revision 21-Jan-2000 #text\_change 21-Jul-2000  
 C/Accession: T45025  
 R/Deaseyn, J.L.; Guyonnet-Duperrat, V.; Porchet, N.; Aubert, J.P.; Laine, A.  
 J. Biol. Chem. 272, 3168-3178, 1997  
 A/Title: Human mucin gene MUC5B, the 10.7 kb large central exon encodes various alternat  
 A/Reference number: Z22899; MUID:97166151; PMID:9013550  
 A/Accession: T45025  
 A/Status: preliminary; translated from GB/EMBL/DBJ  
 A/Molecule type: DNA  
 A/Residues: 1-3570 <DESeq>  
 A/Cross-references: EMBL:Z27496; NID:G1834502; PIDN:CAA96577.1; PID:G1834503  
 A/Experimental source: Placenta  
 C/Genetics:  
 A/Gene: MUC5B

Alignment Scores:  
 Pred. No.: 0.00725 Length: 3570  
 Score: 124.50 Matches: 105  
 Percent Similarity: 31.54% Conservative: 36  
 Best Local Similarity: 23.49% Mismatches: 146  
 Query Match: 4.67% Indels: 160  
 DB: 2 Gaps: 22

US-10-020-441-1 (1-1477) x T45025 (1-3570)

QY 1455 TCCACATCTTTCGCGACACTGCGATATCGACATGATATCGGAGAGTTTGCGAGTGTG 1396  
 Db 695 SerThrValLeuThrThrThrAlaThrThr-----GlyAlaThrGlySerVal 711  
 QY 1395 ---ACACCATGATGATAG----- 1381  
 Db 712 AlaThrProSerSerThrProGlyThrAlaHisThrThrLysValLeuThrThrThr 731  
 QY 1380 -----ACACCGAGTACGTGAGCTTGAGGAGTGGAGAGAGATGCA--- 1339  
 Db 732 ThrGlyPheThrAlaThrProSerSerSerProGlyArg---AlaArgThrLeuProVal 750  
 QY 1338 -----CCACTGCACCTGAGCGCGGATGCTGCCAAATTTGCA 1300  
 Db 751 TrpIleSerThrThrThrThrProThrThrArgGlySerThrValThrProSerSerIle 770

QY 1239 CCTGCTTTCACACATATAGATGCGATGCGATGCGGATTCGATTAATAGTGGCT 1240  
 Db 771 ProGlyThrThrHis-----ThrProThrValLeuThrThrThrThrValAla 788  
 QY 1239 CGACCTGTGGTGCACAAACAAACACACACACAGTACAGT----- 1198  
 Db 789 ThrGlySerMetAlaThrProSerSerSerThrGlnThrSerGlyThrProProSerLeu 808  
 QY 1197 ACAACAGCTGACGATGAGATGAGAAAGAACAAACACACAGGAAACAAACAAACA--- 1141  
 Db 809 ThrThrThrAlaThrThrIleThrAlaThrGlySerThrThrAsnProSerSerThrPro 828  
 QY 1140 ---CACAAGTGAATTTCCAAATATCACTGCGACAAATATCCAC----- 1099  
 Db 829 GlyThrThrProIleProProValLeuThrThrThrAlaThrThrProAlaAlaThrSer 848  
 QY 1098 -----CTTACACTACCCACACACACAAATGTCCACATACA 1063  
 Db 849 SerThrValThrProSerSerAlaLeuGlyThrThrHisThrProProVal----- 865  
 QY 1062 TCACATGTCACATATTCAGACATCAAGTCACTCTCCATCTCCACTCCAAACACA 1003  
 Db 866 -----ProAsn-ThrThrAlaThrThrHis 873  
 QY 1002 TACTCAACA-----CATTGACCCCTACTTCATGATGACACACACACTTCATCGT 952  
 Db 873 eGlyArgSerLeuSerProSerSerProHisThrValCysThrAlaThrThrSerAlaThr 893  
 QY 951 T-----TGCATGATATTCATGAGTCTCTCTCCATGAGTGTCAATCATCC-- 900  
 Db 893 rSerGlyIleLeuGlyThrThrHisIleThrGluProSerThrGlyThrSerHisThrPr 913  
 QY 899 -----ATCAGATCAATTCGATCCGATCCCAATGCAATGCACTCAACACACCTC 856  
 Db 913 oAlaAlaThrThrGlyThrThrGlnHisSerThrProAla----- 926  
 QY 855 ACATCACTGACCTCACTCACTCAACACAAATGACCGAGATGGAATCACTGTCTACTGA 796  
 Db 927 -----LeuSerSerProHisPro-----SerSerArgThrHis 937  
 QY 795 ACTGTCAACAAATGTCACAAACACAGTTGTTATTATCAAGGCTGTGACCAATATACAC 736  
 Db 937 rGlnSerProProSerProGlyThrThrThr-----ProGlyHisThrHis 952  
 QY 735 CGGATCAGTGAACAGACAGACACACAAACAGATACACAGACACAGAAATATGATGATG 676  
 Db 952 rAlaThrSerArgThrThrAlaThrAlaThrProSerLysThrArgThrSerThr----- 970  
 QY 675 ACATCTGACTGACAGTGAATGATGTCGTTTATTCACCAACACACTCTTCTCAATAT 616  
 Db 971 -----LeuLeuProSerGlnProThr----- 977  
 QY 615 TCCACCATTTTACGTACGATCGATACCGTCTGTTATTCCTTCATTAACCAACATGAA 556  
 Db 978 -----SerAlaProIleThrThrValVa 985  
 QY 555 AAC----- 552  
 Db 985 lThrMetGlyCysGluProGlnCysAlaThrSerGluTrpLeuAspTyrSerTyrProMe 1005  
 QY 551 -CCAGTTTCAGATGAGCGGATATCCGAT-----GGCTGTGCGACAGCT 511  
 Db 1005 cProGlyProSerGlyGlyAspPheAspThrTyrSerAsnIleArgAlaAlaGlyGlyAl 1025  
 QY 510 GATCACTCTGATTCACCTC-----TGTAACTTGACATTTGACCCAGACATTAACAT 460  
 Db 1025 aValCysGluGlnProLeuGlyLeuGlyCysArgAlaGlnAlaGlnProGlyValProLe 1045  
 QY 459 TGCATATCAATCACTCACTGAGTGTCTGTGCGATT-----GCTCTCTGTCTGTGCGGA 406  
 Db 1045 uArgGluLeuGlyGlnValValGlyCysSerLeuAspPheGlyLeuValCysArgAsnAr 1065  
 QY 405 CATCAAGAGGCGATGTAT 387



QY 382 CCAGATACATGCCCTTGTATGTCGACACAGAGAGACAATCGACAGACATC 441  
Db 117 ----- 117  
QY 442 AGTGATTCGATATTGCAATGTTAATGCTGGCTCAATGCTCACTTACAGAGTGAATC 501  
Db 118 IlegialaAspIleAlaLeuLeuGluGluProValAsnValSerSerHisVal 137  
QY 502 AGAGATGACGCTCGCACAGCATCGCATCCGCCACCGCAACTGGTGTTCATT 561  
Db 138 HistHValThrLeuProProAlaSerGluThrPheProProGlyMetProCysTrpVal 157  
QY 562 GTTGTTATGAGAGGATGATTAACGAC 588  
Db 158 ThrGlyTrpGlyAspValAspAsnAsp 166  
RESULT 6  
C35863  
Crypsase (EC 3.4.21.59) III precursor - human  
C/Species: Homo sapiens (man)  
C/Date: 03-Feb-1994 #sequence revision 03-Feb-1994 #text\_change 15-Jun-2001  
C/Accession: C35863; A35863; A38893  
R/Vanderlilce, P.; Ballinger, S.M.; Tam, E.K.; Goldstein, S.M.; Craik, C.S.; Caughey, G.  
Proc. Natl. Acad. Sci. U.S.A. 87, 3811-3815, 1990  
A/Title: Human mast cell crypsase: multiple cDNAs and genes reveal a multigene serine pr  
A/Reference number: A35863; MUID:90251647; PMID:2187193  
A/Accession: C35863  
A/Molecule type: mRNA  
A/Residues: 9-275 <VAN>  
A/Accession: A35863  
A/Molecule type: DNA  
A/Residues: 1-9 <VA2>  
A/Cross-references: GB:M33494; NID:g3927804; PID:AA083172.1; PID:g339977  
A/Note: the first nine residues of this sequence are inferred from genomic DNA of crypca  
R/Vanderlilce, P.  
submitted to GenBank, April 1990  
A/Reference number: A38893  
A/Accession: A38893  
A/Molecule type: mRNA  
A/Residues: 9-131, 'K, 132-275 <VA3>  
A/Cross-references: GB:M33493; NID:g339984; PID:AAA36780.1; PID:g339985  
C/Superfamily: trypsin; trypsin homology  
C/Keywords: hydrolase; serine proteinase; zymogen  
F/1-21/Domain: signal sequence #status predicted <SIG>  
F/22-30/Domain: activation peptide #status predicted <ACT>  
F/31-275/Product: trypsin I #status predicted <MAT>  
F/31-267/Domain: trypsin homology <TRY>  
F/74,121,224/Active site: His, Asp, Ser #status predicted  
Alignment Scores:  
Pred. No.: 0.00781 Length: 275  
Score: 123.00 Matches: 38  
Percent Similarity: 35.50% Conservative: 22  
Best Local Similarity: 22.49% Mismatches: 45  
Query Match: 4.58% Indels: 64  
Gaps: 3  
US-10-020-441-1 (1-1477) x C35863 (1-275)  
QY 91 TGTACAGSTTCACTAGTCTCAAGAGACGACTCACTGTCATTTGTT----- 144  
Db 59 CysGlyGlySerLeuIleHisProGlnTrpValLeuThrAlaHisCysValGlyPro 78  
QY 145 ---TGTCAACATGCGCAGTGAATTCGGGTAAAGATCGAAGAACACATTTGTTCTCA 201  
Db 79 AspValIlyAspLeuAlaIleLeuArgValGlnLeuArgGluGlnHisLeuIlyTrpGln 98  
QY 202 TGTATTCATTTATTCACAGACAGAGTGTGGCTGCGTGGCTGGCTTTGTGTGTATGC 261  
Db 99 AspGln----- 100  
QY 262 AGTGGTTTGCATAGATTTTATGCTATATACCTTGTGTTATTCAGGTTTCATTCTC 321

Db 101 -----LeuLeuProValSerArgIleIleValHisProGlnPheTrp 114  
QY 332 ACAGTGAAGATGGGACCAACAGGATCCATACCAACGCTGAGTTAAGTTGCA 381  
Db 115 ThrAlaGln----- 117  
QY 382 CCAGATACATGCCCTTGTATGTCGACACAGAGAGACAATCGACAGACATC 441  
Db 117 ----- 117  
QY 442 AGTGATTCGATATTGCAATGTTAATGCTGGCTCAATGCTCACTTACAGAGTGAATC 501  
Db 118 IlegialaAspIleAlaLeuLeuGluGluProValAsnValSerSerHisVal 137  
QY 502 AGAGATGACGCTCGCACAGCATCGCATCCGCCACCGCAACTGGTGTTCATT 561  
Db 138 HistHValThrLeuProProAlaSerGluThrPheProProGlyMetProCysTrpVal 157  
QY 562 GTTGTTATGAGAGGATGATTAACGAC 588  
Db 158 ThrGlyTrpGlyAspValAspAsnAsp 166  
RESULT 7  
B46629  
mucin 6, gastric (3-repeat clone) - human (fragment)  
N/Alternate names: MUC6  
C/Species: Homo sapiens (man)  
C/Date: 21-Sep-1993 #sequence revision 18-Nov-1994 #text\_change 05-Nov-1999  
C/Accession: B46629  
R/Toribara, N.W.; Robertson, A.M.; Ho, S.B.; Kuo, W.L.; Gum, E.; Hicks, J.W.; J  
J. Biol. Chem. 268, 5879-5885, 1993  
A/Title: Human gastric mucin. Identification of a unique species by expression cloning.  
A/Reference number: A46629; MUID:91194895; PMID:7680650  
A/Accession: B46629  
A/Status: Preliminary  
A/Molecule type: mRNA  
A/Residues: 1-505 <TOR>  
A/Cross-references: GB:I07518; NID:g292045; PID:AA061945.1; PID:g292046  
A/Experimental source: Stomach  
A/Note: Sequence extracted from NCBI backbone (NCB1N:128397, NCB1P:128399)  
C/Genetics:  
A/Gene: GDB:MUC6  
A/Cross-references: GDB:I34734; OMIM:158374  
A/Map position: 11p15.5-11p15.5  
C/Keywords: glycoprotein  
Alignment Scores:  
Pred. No.: 0.0113 Length: 505  
Score: 121.50 Matches: 96  
Percent Similarity: 34.57% Conservative: 34  
Best Local Similarity: 25.53% Mismatches: 150  
Query Match: 4.56% Indels: 96  
Gaps: 20  
US-10-020-441-1 (1-1477) x B46629 (1-505)  
QY 1455 TCACAATTTAGCCACACTGGCATCTGCAATGATATGCGAAGSTTTGGCACTGTG 1396  
Db 68 ThrAsnIlyThrProThrSerLeuHisSerHisThrSerSerThrHisProGluVal 87  
QY 1395 ACACCATGATGATGACG----- 1378  
Db 88 AlaProThrSerThrThrThrIleThrProAsnProThrSerThrArgThrArgThrPro 107  
QY 1377 ---CCGACTACTGACCTTGAAGGATGGAGAGAGTCCACGACTG-----TCA 1330  
Db 108 ValAlaHisThrAsnSerAlaThrSerSerArgProProProPheThrHisSer 127  
QY 1339 CCTGAGCGGTAAGTGTGCAAAATTCGACCTGCTTACACAAATPAGATTGGCATTTG 1270  
Db 128 ProProThrGlySerSerPro-----PheSerSerThrGly---ProMet 141

```
OY 1269 GTCCGATGTCGGCATTTCATTAATAGTCTGCACCCCTGTGTGCAAAACAACACCACAA 1210
    |||
Db 142 ThrAlaThrSerPheLys----- 147
OY 1209 CAACAGTACAGTACAAACAGTACAGTACAGTACAGTACAGTACAGTACAGTACAGTAC 1150
    :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 148 -----ThrThrThrThrThrThrThrThrThrThrThrThrThrThrThrPro----- 162
OY 1149 AACAAACACAGACAGTATTCGAAATATGCAATGAGCAGCAATATCCACCCCTTACACT 1090
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 163 ---LeuThrHis-----ValProProPheSerThrSerLeuValThrProle----- 177
OY 1089 ACCCAGACACAAATA-----GTCCAGATACAAATCAGATGCAATATTCGACAGATCA 1036
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 178 ThrHisThrValIleThrProThrHisProGlnMetSerThrSerAlaTyrIleHisSer 197
OY 1035 AATGACTCTCCATCTGCACAGCTCTGCAACACATACCTAAACACATGACCCCTTCTTC 976
    :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 198 ThrProThrGlyThrIleAlaSerProThrThrValIleAlaThrArgSerThrTyrThr 217
OY 975 -----ATGTACACACACACACATTCCTGCTTGCATAGATATTTTCATGAGTCTTCTGCC 922
    :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 218 AlaProLeuMetThrAlaThrThrSerArgIleSerGlnAlaHisSerSerIleSerThr 237
OY 921 TCCAAAGTGTCTCATCATCCAT-----CACATGACATTCATGCC--ATCCAA 874
    :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 238 AlaIysThrSerThrSerLeuHisSerHisAlaSerSerThrHisHisProGlnValThr 257
OY 873 TCTACCTCAACT-----CACCTCACATCCAGT----- 847
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 258 ProThrSerThrThrAsnValThrProLysSerThrSerArgAspThrSerThrProVal 277
OY 846 ---CACCTCACCTCACACACATGACCGAGAT-----GGAATGATCTCT-- 804
    :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 278 ThrHisThrThrSerAlaThr--SerSerArgProProThrProIleThrThrHisSerSe 297
OY 803 -----GCTACTGAACTGTACAAAGTCATCCAAACACAGATGTTATTAATCAAGGC 754
    :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 297 rProThrArgSerSerProLeuSerSerThrArgIlyProMetThrAlaThrSerIleLysTh 317
OY 753 TCTGACCAATACAAACCGGATCAGGTAGACAGACAGACAGACAAACAGATCA----- 699
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 317 rThrThrThrTyrProThrProSerHisProGlnThrThrLeuThrThrHisValProPr 337
OY 698 ----CAGACACAGATACTGACATGACATGACATGACATGACATGACATGACATGACAT 643
    :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 337 oPheSerThrSerValThrProSerThrHisThr-----ValIleThrProth 354
OY 642 TCACCAACAACCTCATTTCTTCAAATATCCACCATTTCTTACGTGACGAGATCAGGTGTT 583
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 354 rHisAlaGlnMetSer-----ThrSerAlaSerIle 364
OY 582 ATCATCCCTTCCA-----TAAACAACAATGAAAAACACAGTTCACAGTGGCGGGAT 532
    :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 364 eHisSerThrProThrGlyThrValProProLeuThrThrArgMetProThrIleLysThr 384
OY 531 ATCCGATGCTGTGCGACAGATGATCTGATTCATTCATCTGTGTAAG 486
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 384 rArgThrGlyProProMetThrGlyThrIleIleGlnThrSerLys 399

RESULT 8
T13596
trypsin homolog - fruit fly (Drosophila melanogaster)
N:Alternate names: protein EG:80H7.1
C:Species: Drosophila melanogaster
C>Date: 13-Aug-1999 #sequence_revision 13-Aug-1999 #text_change 17-Nov-2000
C:Accession: T13596
R:Benos, P.
A:Submitted to the EMBL Data Library, April 1999
A:Description: Sequencing the distal X chromosome of Drosophila melanogaster.
A:Reference number: Z17667
A:Accession: T13596
A>Status: preliminary; translated from GB/EMBL/DBJ
,
```

```
A:Molecule type: DNA
A:Residues: 1-281 <BEN>
A:Cross-references: EMBL:AL031027; PIDN:CA19843.2
C:Genetics:
A:Cross-references: FlyBase:FBgn0000481
A:Introns: 159/3
A>Note: EG:80H7.1
C:Superfamily: trypsin; trypsin homology

Alignment Scores:
Pred. No.: 0.0119 Length: 281
Score: 121.00 Matches: 51
Percent Similarity: 34.55% Conservative: 25
Best Local Similarity: 23.18% Mismatch: 70
Query Match: 4.50% Indels: 74
DB: Gaps: 7

US-10-020-441-1 (1-1477) x T13596 (1-281)
OY 49 GAATTCGGTTCATGCAATCTTA-----ACGACAGAGAGAAACAATGTATACAGST 99
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 37 GluPheProPheValValSerLeuArgArgAlaLysSerGlyArgHisSerCysGlyAla 56
OY 100 TCACTAGTCTCAACGAGACAGTACTGACAGTGTGCTCATTTGTTGTTGCTCAGCATGGCA 159
    :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 57 ThrLeuLeuAsnProTyrTrpValLeuThrAlaAlaHisCysValArgGlySerSerPro 76
OY 160 GTGATTCGGGTAAAGATGACAGTCAAAACATGTTGCTCAATGTATGATGATTATTC 219
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 77 Glu----- 77
OY 220 CAGACAGTGTTCGTCGTGCGTGGCTTGTGTGTGTATGACAGTGGTTGCATGAT 279
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 77 ----- 77
OY 280 TTAATGCTATTAATGCTTATTTGATTTTATTCAGTTTATTCACATGACAGTAATGGCAGC 339
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 78 -----GlnLeuAspLeuGlnTyrGlySerGlnMetLeuAlaArgAsnSerSer 93
OY 340 CAACAAGGATCCATCACCAACCGCTGAGTAAAGTGGGACCGAGATACATGCGCTCT 399
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 94 GlnValAla-----ArgValAlaAlaIlePheValHisProGlyTyrGluProGlu 110
OY 400 TGTATGTGGGACGACAGAGAGACCAATCGCAGACACTCATGATGATTCGATATTCGA 459
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 111 AspLysTyrValAsn-----AspIleLeu 118
OY 460 ATTGTAATGCTGGCTCAAAATGCTCACTTACAGAGTGAATACAGATGATGATGCTGCCA 519
    :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 119 LeuLeuGlnLeuAlaGlnSerValAlaLeuSerLysPheValGlnProValArgLeuPro 138
OY 520 CAGCATGGATATCCCGCCACCTGGAACCTGATGTTTCATGTTGTTGTTATGAGGGAT 579
    :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 139 GluProArgGlnValThrProGlyAsnAlaSerAlaValLeuAlaGlyTyrProLysLeuAsn 158
OY 580 GATAACGACCGGTGATCCGTCAAGTAAAGATGTTGGAATA-----TTGAAGAA 627
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 159 -----AlaThrGlyGlyValValGlnGlnIleGlnIleGlnLys 170
OY 628 GTGAGTGTGTGTAATAACGACATGACTCAGTGTGCTCAGTCAATGTCAGTCACTTAT 687
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 171 ValLysLeuGlnValPheSerAsp---ThrGluCysSerGluArgHisGlnThrTyrLeu 189

RESULT 9
S53362
mucin 5AC (clone JER47) - human (fragment)
C:Species: Homo sapiens (man)
C>Date: 28-Oct-1996 #sequence_revision 07-Feb-1997 #text_change 08-Oct-1999
R:Guyonnet-Duprat, V.; Audie, J.P.; Debailleul, V.; Laine, A.; Buistine, M.P.; Galliegue
Biochem. J. 305, 211-219, 1995
A:Title: Characterization of the human mucin gene MUC5AC: a consensus cysteine-rich dom
A:Reference number: S53361; MUID:95126907; PMID:7826332
```



A:Accession: S53362  
 A:Status: nucleic acid sequence not shown  
 A:Molecule type: mRNA  
 A:Residues: 1-477 <GUV>  
 A:Cross-references: EMBL:Z34277  
 A:Experimental source: clone JER47  
 R:Porter, N.  
 Submitted to the EMBL Data Library, June 1994  
 A:Reference number: S71065  
 A:Accession: S71065  
 A:Molecule type: mRNA  
 A:Residues: 1-211, S', 213-224, 'AR', 227-259, 'S', 261-477 <POR>  
 A:Cross-references: EMBL:Z34277; NID:G563374; PIDN:CAA4031.1; PID:G563375  
 A:Experimental source: clone JER47  
 C:Genetics:  
 A:Gene: GDB:MUC5AC  
 A:Cross-references: GDB:454136; OMIM:158373  
 A:Map position: 11p15.5-11p15.5  
 C:Keywords: glycoprotein; tandem repeat

Alignment Scores:  
 Pred. No.: 0.0154 Length: 477  
 Score: 120.00 Matches: 103  
 Percent Similarity: 34.85% Conservative: 35  
 Best Local Similarity: 26.01% Mismatches: 153  
 Query Match: 4.50% Gaps: 105  
 DB: 2 13

US-10-020-441-1 (1-1477) x S53362 (1-477)

QY 1444 AGCCACATCGGATATCGACATATCGGGAGGTTGGAGTGTAC----- 1394  
 DB 75 SerHisProGluValSerIleGluHisLeuGlyGlnValGlnCysSerArgGluGlu 94  
 QY 1393 -----ACCATGATTCGAC 1379  
 DB 95 GlyLeuValCysArgAsnGlnAspGlnGlnGlyProPheIleuSerCysLeuAsnTrpGlu 114  
 QY 1378 ACCGATGACTGG-----ACCTGAAGGAGTGGAGAGAGGTCACCACTGTCACC 1328  
 DB 115 ValArgValLeuCysGluThrProGlyGlyCysProValThrSerThrProValThr 134  
 QY 1337 TGGAGCGGGTAACTGCCGAAATTCGACCTGCTTCACACATATGATGGCATTTGCT 1268  
 DB 135 AlaProSerThrProSer-----Gly 141  
 QY 1267 CGCATGCGGATTCATATATGTCGCTGACCTGTGTGCAACAAACACACACACA 1208  
 DB 142 ArgAlaThrSerProThrGlnSerThrSerSer---TrpGlnIleSerArgThrThrThr 160  
 QY 1207 ACAGTACGTACACACAGTACAGTACAGTACAGTACAGTACAGTACAGTACAGTAC 1148  
 DB 161 LeuVal-ThrThrSerThrThrSerThr-----ProGlnThrSerThrThrSerAl 177  
 QY 1147 CAAACACACAGTAAATTCGAAATATCACTGGACATATCCACCTTTACACTAC 1088  
 DB 177 AProThrThrSerThrThrSerAlaProThrThrThrSerAla---ProThrThr 196  
 QY 1087 CCACACACATATGTCACATATCAATCATCATGTCATATATTCACGATCAAGTCACTC 1028  
 DB 196 rSerThrThr-----SerThrProG1 203  
 QY 1027 TCCATCTACACACTGCCAAACACATATCAACATGACCCCTACTTCACTGACAC 968  
 DB 203 nThrSerIleSerSerAlaProThrThrThrThrSerAlaPro----- 218  
 QY 967 AACACACTTCATCGTTGTCATAGATATTCATGAGTCTTCTCTCCCAAGTGTCTCA 908  
 DB 219 -----ThrSerThrIleSerThr 225  
 QY 907 ATCATCCATCATCATCATTCATCCATCCATCCATCCATCCATCCATCCATCCATCCAG 848  
 DB 225 GAlaThrThrSerIleIleSerAlaPro-----ThrThrSerThr-ThrSerPheProT 243

QY 847 TCACCTCACTCAACCAACACATGACCGAGATGGAATCACTCTCTACTGAACTGCAC 788  
 DB 243 hrThrSerThrThrSerAlaThrThrThrSerThrThrSerAlaProThrThrThrThr 263  
 QY 787 AAGTCATCCAAACACAGTGTATATTCATGAGCTCTGACCATATCAACACCGATCAG 728  
 DB 263 hrSerThrProGlnThrSerIleThrSerAlaThrSerSerThrThrSerGlySerG 283  
 QY 727 GTAGACAGACACACACAAACAGTATACAGACACACAGATATCACTGATCACTGCA 668  
 DB 283 lyThrThrProSerProValThrThrThrSerThrAlaSerValSerIleThrSer-- 301  
 QY 667 CTGACGACTGATGATGCTGTTATTCACCAACACTCACTTCTTCATATATCCACCAT 608  
 DB 302 -----ThrSerHisValSerValSerIleThrThrThrSerGlnProValThrArgAs 319  
 QY 607 TCTTACGTACCGATCAG-----GTCTTATCATCCCTTC----- 571  
 DB 319 pCysHisProArgCysThrThrThrThrThrThrThrThrThrThrThrThrThrThr 339  
 QY 570 -----ATAACCAACATGAAACACACAGTTC 545  
 DB 339 ohIleGlyLeuPheGlyGluThrThrThrThrThrThrThrThrThrThrThrThrThr 359  
 QY 544 CAGTGGCGGATATCCGATGCTGTCGACACTGATCACTGATTCACACTGTAAGT 485  
 DB 359 gArgProGluGluIleThrThrThrThrThrThrThrThrThrThrThrThrThrThr 379  
 QY 484 TGACCATTTGACCCAGCATTCATCAATTCGATTCGATTCGATGATGCTGTGGCATTTG 425  
 DB 379 eGluHisLeuGlyGlnValGlnCysSerHisGlu-----Glu-G 393  
 QY 424 GTCTCTCTGTGTGTCCGACATCAAGAGGGCATGTAT 387  
 DB 393 lYleuValCysArgAsnGlnAspGlnGlnGlyProPhe 405

RESULT 10  
 A29934  
 pancreatic elastase (EC 3.4.21.36) IIITA precursor - human  
 N:Alternate names: protease E  
 C:Species: Homo sapiens (man)  
 C:Date: 15-Dec-1988 #sequence\_revision 15-Dec-1988 #text\_change 20-Jun-2000  
 A:Accession: A29934; JX0045  
 R:Tani, T.; Ohsumi, J.; Mita, K.; Takiguchi, Y.  
 J. Biol. Chem. 263, 1231-1239, 1988  
 A:Title: Identification of a novel class of elastase isozyme, human pancreatic elastase  
 A:Reference number: A92664; MUID:88087253; PMID:2826474  
 A:Accession: A29934  
 A:Molecule type: DNA  
 A:Residues: 1-270 <TAN>  
 A:Cross-references: GB:U03516  
 R:Shirai, Y.; Takemura, K.; Yoshida, H.; Sato, Y.; Iijima, H.; Shimada, Y.; Miyama, Y.  
 J. Biochem. 104, 259-264, 1988  
 A:Title: Molecular cloning of complementary DNA encoding one of the human pancreatic pr  
 A:Reference number: JX0045; MUID:89034017; PMID:2460440  
 A:Accession: JX0045  
 A:Molecule type: mRNA  
 A:Residues: 1-240, 'G', 242-270 <SHI>  
 A:Cross-references: GB:D00306; NID:G220013; PIDN:BA00212.1; PID:G220014  
 C:Comment: This enzyme is an alanine-specific serine proteinase that has little elastol  
 C:Genetics:  
 A:Introns: 15/1; 43/3; 76/2; 121/2; 167/1; 214/3; 265/3  
 C:Superfamily: trypsin; trypsin homology  
 C:Keywords: hydrolase; pancreas; serine proteinase  
 F:1-15/Domain: signal sequence #status predicted <SIG>  
 F:16-28/Domain: propeptide #status predicted <PRO>  
 F:29-270/Product: pancreatic elastase IIITA #status experimental <MAT>  
 F:29-265/Domain: trypsin homology <TRY>  
 F:73,123,217/Active site: His, Asp, Ser #status predicted

Alignment Scores:  
 Pred. No.: 0.0221 Length: 270

Score:	118.00	Matches:	43
Percent Similarity:	34.3%	Conservative:	26
Best Local Similarity:	21.3%	Mismatches:	60
Query Match:	4.3%	Indels:	72
DB:	2	Gaps:	4

US-10-020-441-1 (1-1477) X A29934 (1-270)

OY		25	GGGAAACCTGTGCAACACCGAGACT-----	48
Dd		19	GlyProPheSerHisSerSerArgValAlaHISglutInuAspAlaValProTyr	38
OY		49	GAATTCGCCGTTTCATCGCATTCCTTAACGACAGAGAACA-----ATTGCT	93
Dd		39	SerTrpProTrpGlnValSerLeuGlnTyrGluIysSerGlySerPheTyrIleThrCys	58
OY		94	ACAGSTTCACTAAGTCTCACACAGACAGACTACTCCACAGCTGCTCATTTGTGTTGCTACCA	153
Dd		59	GlyGlySerLeuIleAlaProAspTrpValValThrAlaGlyHisCysIleSerArgAsp	78
OY		154	TTCGCAGTGATTCGGGTAAAGATGCACCTGAAAACACATTGTGCTCAATGTAATTCGATTG	213
Dd		79	Leu-----	79
OY		214	ATTTCACAGACAGATGTTTGCCTGTCGCTGCTGTGTGTGTATGCAGTGGGTTTTGCA	273
Dd		79	-----	79
OY		274	TAGATTTTAATGTCTATATACTGTGGTCTTATTACAGTTTCATTTCACACTGAGGAAT	333
Dd		80	-----ThrTyrGlnAlaValLeuGlyGluTyrAsnLeuAlaValLysGlu	94
OY		334	GGCGACCAACAAGGCATCCATACCACCAACCGCTGTGAAGTTAAGTGCACAGACGATACATG	393
Dd		95	GlyProGluGlnValIleIleProIleAsnSerIuGluIuePheValHisProLeuTrpAsn	114
OY		394	CCCTCTTGTATGTGGCGACGACAGAGAGACCAATTCGACACAGACATCACTGATGATTCGAT	453
Dd		115	ArgSerCysValAlaLys-----GlyAsnAsp	123
OY		454	ATTGCAATTGTAAATCTGTGGCTCAATATGCTCAACTTACAGTAGTGGAATCAGAGTATCAGT	513
Dd		124	IleAlaLeuIleLysLeuSerArgSerArgIleGlnIleuGlyAspAlaValGlnLeuAlaSer	143
OY		514	CTGCCACACCATTCGATATCCCGCCACCTGGAACCTGGATGTTTCATTGTGTGTTATGGA	573
Dd		144	LeuProProAlaGlyAspIleLeuPheProAsnLysThrProCysTyrIleThrGlyTrpGly	163
OY		574	AGG 576	
Dd		164	Arg 164	

RESULT 11  
S56160  
mast cell tryptase precursor - Mongolian jird  
C|Species: Meriones unguiculatus (Mongolian jird)  
C|Date: 27-Oct-1995 #sequence\_revision 03-Nov-1995 #text\_change 20-Jun-2000  
C|Accession: S56160  
R|Makukuno, Y.; Ide, H.; Itoh, H.; Tomita, M.; Kobayashi, T.; Maruyama, H.; Horii, Y.; N  
Biochem. J. 309, 921-926, 1995  
A|Title: Cloning of the cDNA encoding mast cell tryptase of Mongolian gerbil, Meriones u  
A|Reference number: S56160; MUID:95366971; PMID:7639711  
A|Accession: S56160  
A|Status: preliminary  
A|Molecule type: mRNA  
A|Residues: 1,270 <MNP>  
A|Cross-references: EMBL:D31789; NID:G517122; PIDN:BA06598.1; PID:G517123  
C|Superfamily: trypsin; trypsin homology  
F|26-262/Domain: trypsin homology <TRY>

Alignment Scores:	
Pred. No.:	0.0272
Score:	117.00
Length:	270
Matches:	42

Percent Similarity:	34.23%	Conservative:	2
Best Local Similarity:	23.20%	Mismatches:	3
Query Match:	4.36%	Indels:	8
DB:	2	Gaps:	5

US-10-020-441-1 (1-1477) X S56160 (1-270)

[illegible]

B35863 12  
 trypsinase (EC 3.4.21.59) II precursor - human  
 N:Alternate names: trypsinase beta  
 C:Species: Homo sapiens (man)  
 C:Date: 03-Feb-1994 #sequence revision 03-Feb-1994 #text\_change 21-Jul-2000  
 C:Accession: B35863; A37193; I59473  
 R:Vanderlinden, P.; Ballinger, S.M.; Tam, E.K.; Goldstein, S.M.; Craik, C.S.; Caughey, G.  
 Proc. Natl. Acad. Sci. U.S.A. 87, 3811-3815, 1990  
 A:Title: Human mast cell trypsinase: multiple cDNAs and genes reveal a multigene serine p  
 A:Reference number: A35863; MUID:90251647; PMID:2187193  
 A:Accession: B35863  
 A:Molecule type: mRNA; DNA  
 A:Residues: 1-275 <MAN>  
 A:Cross-references: GB:M33497; NID:G339982; PID:AAA36779.1; PID:G339983  
 A:Note: Residues 2-275 are derived from mRNA; residue one was inferred from the genomic  
 R:Miller, J.S.; Moxley, G.; Schwartz, L.B.  
 J. Clin. Invest. 86, 864-870, 1990  
 A:Title: Cloning and characterization of a second complementary DNA for human trypsinase  
 A:Reference number: A37193; MUID:90369005; PMID:2203827  
 A:Accession: A37193  
 A:Molecule type: mRNA  
 A:Residues: 1-275 <MIT>  
 A:Cross-references: GB:M37488; NID:G179583; PID:AAA51843.1; PID:G179584  
 R:Blom, T.; Hellman, L.  
 Scand. J. Immunol. 37, 203-208, 1993

A:Title: Characterization of a tryptase mRNA expressed in the human basophil cell line K  
 A:Reference number: 159473; MUID:93166209; PMID:8434231  
 A:Accession: 159473  
 A:Status: preliminary; translated from GB/EMBL/DBJ  
 A:Molecule type: mRNA  
 A:Residues: 1-275 <RES>  
 A:Cross-references: GB:S55551; NID:9265666; PIDN:AAD13876.1; PID:94261576  
 A:Experimental source: basophil cell line KU812  
 C:Genetics:  
 A:Gene: GDB:TPS1  
 A:Cross-references: GDB:125890; OMIM:191080  
 A:Map position: 16pter-16qter  
 C:Superfamily: trypsin; trypsin homology  
 C:Keywords: hydrolase; serine proteinase; zymogen  
 F:1-21/Domain: signal sequence #status predicted <SIG>  
 F:22-30/Domain: activation peptide #status predicted <ACT>  
 F:31-275/Product: trypsin I #status predicted <MAT>  
 F:31-267/Domain: trypsin homology <TRI>  
 F:74,121,224/Active site: His, Asp, Ser #status predicted

Alignment Scores:

Pred. No.:	0.0273	Length:	275
Score:	117.00	Matches:	37
Percent Similarity:	34.91%	Conservative:	22
Best Local Similarity:	21.89%	Mismatches:	46
Query Match:	4.36%	Indels:	64
DB:	2	Gaps:	3

US-10-020-441-1 (1-1477) x B35863 (1-275)

```

QY 91 TGTACAGSTTCCTGATCTCAACAGAGAGTACTCAGCTGCTGATTTGTT----- 144
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 59 CysGlyGlySerLeuIleHisProGlnTrpValLeuThrAlaHisCysValGlyPro 78
QY 145 ---TCTCAACCTTGGCAGCTGATTTGGGTAAGATCCGCTGAACACATTTGGTCTCA 201
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 79 AspValIlyAspLeuAlaAlaLeuArgValGlnLeuArgGlnGlnHisLeuTyrTyrGln 98
QY 202 TGTATCATGATTGTTCAACAGACAGTGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 261
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 99 AepGln----- 100
QY 262 AGTGGTTTGCATGATTTTAATGCTATATATCTGCTTATTTTCAAGTTTCATTTCTC 321
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 101 -----LeuLeuProValSerArgIleIleValHisProGlnPheTyr 114
QY 322 ACACGAGGAATGGCGACCAACAGGATCCATCACCAACGCTGGAGTTAAGTGCA 381
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 115 ThrAlaGln----- 117
QY 382 CCAGATACATGCTCTTGTATGTCGGACAGACAGAGAGACCAATGCGACACATC 441
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 117 ----- 117
QY 442 AGTGATTCGATATTGCAATTGTAATGCTGCTCAATGCTCACTTACAGAGTGAATC 501
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 118 IleGlyAlaAspIleAlaLeuLeuGlnLeuGlnProValIleValSerHisVal 137
QY 502 AGAGTCATCAGTCTCCACACAGCATCGATATCCGCCACTGGAACCTGGTGTTCATT 561
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 138 HisThrValIleThrLeuProAlaSerGlnThrPheProGlnPheProCysTrpVal 157
QY 562 GTTGGTATGGAAGGATGATTAAGAC 588
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 158 ThrGlyTrpGlyAspValAlaAspAsp 166

```

RESULT 13

S57180  
 probable membrane protein YJR151c - yeast (Saccharomyces cerevisiae)  
 N:Alternate names: protein J2223; serine/threonine-rich protein YJR151c  
 C:Species: Saccharomyces cerevisiae  
 C:Date: 23-Aug-1995 #sequence\_revision 08-Sep-1995 #text\_change 19-Apr-2002  
 C:Accession: S57180

R:Scarcet, T.  
 submitted to the Protein Sequence Database, September 1995  
 A:Reference number: S57180  
 A:Accession: S57180  
 A:Molecule type: DNA  
 A:Residues: 1-1161 <SCA>  
 A:Cross-references: EMBL:Z49651; NID:91015902; PID:91015903; MSPDB:GN00010; MIPS:YJR151  
 C:Genetics:  
 A:Gene: SGD:DNA4; MIPS:YJR151c  
 A:Cross-references: SGD:S0003912  
 A:Map position: 10R  
 C:Keywords: transmembrane protein

Alignment Scores:

Pred. No.:	0.0426	Length:	1161
Score:	115.50	Matches:	62
Percent Similarity:	37.64%	Conservative:	37
Best Local Similarity:	23.57%	Mismatches:	127
Query Match:	4.33%	Indels:	37
DB:	2	Gaps:	3

US-10-020-441-1 (1-1477) x S57180 (1-1161)

```

QY 1270 GGTGGCATGTGGCATTTCCATTATGCTGACCTGTGGTCAACAAACACCA 1211
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 116 GlyIleTyrThrAlaIleProThrSerThrSerThrThrThrThrThrSerThrSer 135
QY 1210 ACAACAGTACAGTCAACAGTACAGTACAGTACAGTACAGTACAGTACAGTACAG 1151
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 136 ThrThrPro-ThrThrThrThrThrThrThrThrThrThrThrThrThrThrThrThr 155
QY 1150 AAACAAACACCAAGTGAATTTGCAAAATATCACTGGCAATATCCACCTTACAC 1091
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 155 rSerThrThrSerThrThrThrThrThrThrThrThrThrThrThrThrThrThrThrThr 175
QY 1090 TACCACCAACCAATAGTGCACATACATACATAGTGCATATATTCAGCATCAAGTCA 1031
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 175 rThrSerThrThr-----ProThrThr 182
QY 1030 CTCCTCATCTCACCACTCTCCAAACACATACTCAACACATGACCCCTCATCATGTA 971
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 182 rSerThrThrSerThrThrProThrThrThrThrThrThrThrThrThrThrThrThrThr 202
QY 970 CACAACACACTTCATCGTTGATAGATATTCATGAGTCTCTCTCCCTCAAGTGTG 911
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 202 ThrThrSerThr-----ThrProThrThrThr 211
QY 910 TCATCATCCATCATCATCATCATTCATCCATCCATCCATCCATCCATCCATCCATC 851
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 211 ThrThrSerThrThrProThrThrThrThrThrThrThrThrThrThrThrThrThrThr 231
QY 850 ACGTCACCTCACCCTCACCAACACATGACGAGATGGAATCATCTGCTACTGAAGCT 791
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 231 rThrThrSerThrThrSerThrThrSerGlnThrThrThrThrThrThrThrThrThrThr 251
QY 790 CACAAGCATCCCAACACAGTGTGTTATTCAGAGGCTGACCAATACCAACACCGGAT 731
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 251 ThrSerThrSerThrThrThrThrThrThrThrThrThrThrThrThrThrThrThrThr 271
QY 730 CAGGTAGACAGACAGACACCAACAGATACAGACAGACAGACAGATACAGACAGATC 671
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 271 ThrAlaProThrThrThrThrThrThrThrThrThrThrThrThrThrThrThrThrThr 290
QY 670 TGACTGACGACTGATGATGCTGTTTATTCACCAACACTGCTTCTTCAATATTCAC 611
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 291 ProThrThrThrThrThrThrThrThrThrThrThrThrThrThrThrThrThrThrThr 310
QY 610 CATCTTACGTCAGGATTCACGCTGTTATCATCTTCAATACCAACATGAAAAAC 551
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 310 ThrSerThrThrThrThrThrThrThrThrThrThrThrThrThrThrThrThrThrThr 330
QY 550 CAGTTCACGAGTGGCGGATATCCATGAGCTGTGGACAGTATCATCTGATTCACCTCT 491
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

Db 330 rThra1a-----SerThraPsiThrThrSerSerVa 341

Qy 490 G 490

Db 341 I 341

RESULT 14

B29934

pancreatic elastase (EC 3.4.21.36) IIIB precursor - human

N:Alternate names: 35k glycoprotein; pancreatic protein P35; protein G32; proteinase E

C:Species: Homo sapiens (man)

C:Date: 15-Dec-1988 #sequence\_revision 15-Dec-1988 #text\_change 22-Jun-1999

C:Accession: B29934; A90516; A27206; A33257; A28932; S04490

J:Taxi: T.; Othumli, J.; Mita, K.; Takiguchi, Y.

J: Biol. Chem. 263, 1231-1239, 1988

A:Title: Identification of a novel class of elastase isozyme, human pancreatic elastase

A:Reference number: A92664; MUID:86087253; PMID:2826474

A:Accession: B29934

A:Molecule type: mRNA

A:Residues: 1-270 <TAN>

A:Cross-references: GB:M18692; NID:9607029; PID:AA58454.1; PID:G182035

R:Shen, W.; Fletcher, T.S.; Laryman, C.

Biochemistry 26, 3447-3452, 1987

A:Title: Primary structure of human pancreatic protease E determined by sequence analysis

A:Reference number: A90516; MUID:88000545; PMID:3477287

A:Accession: A90516

A:Molecule type: mRNA

A:Residues: 'G', 5-63, 'G', 65-78, 'W', 80-118, 'G', 120-163, 'P', 165-270 <SHE>

R:Fletcher, T.S.

submitted to GenBank, August 1987

A:Reference number: A94507

A:Contents: revision to residue 119

A:Accession: A27206

A:Molecule type: DNA

A:Residues: 'G', 5-63, 'G', 65-78, 'W', 80-163, 'P', 165-270 <FLB>

R:Aviles, F.X.; Pascual, R.; Salva, M.; Bonicel, J.; Puligerver, A.

Biochem. Biophys. Res. Commun. 163, 1191-1196, 1989

A:Title: Generation of a subunit III-like protein by autolysis of human and porcine pro

A:Reference number: A33257; MUID:89392022; PMID:2675835

A:Accession: A33257

A:Molecule type: protein

A:Residues: 18-57 <AVI>

R:Guy-Crotte, O.; Barthe, C.; Basso, D.; Fournet, B.; Figarella, C.

Biochem. Biophys. Res. Commun. 156, 318-322, 1988

A:Title: Characterization of two glycoproteins of human pancreatic juice: P35, a truncat

A:Reference number: A28932; MUID:89025862; PMID:3178837

A:Accession: A28932

A:Molecule type: protein

A:Residues: 'X', 32-52, 'X', 54-55, 'XXX', 59-63 <GVY>

R:Mouliard, M.; Kerfelec, B.; Mallet, B.; Chapus, C.

FEBS Lett. 250, 166-170, 1989

A:Title: Identification of a procarboxypeptidase A-truncated protease E binary complex i

A:Reference number: S04999; MUID:89325560; PMID:2753124

A:Accession: S04999

A:Molecule type: protein

A:Residues: 31-50 <MOU>

R:Mendorf, P.; Geyer, R.; Sziegoleit, A.; Linder, D.

FEBS Lett. 249, 275-278, 1989

A:Title: Localization and characterization of the glycosylation site of human pancreatic

A:Reference number: S04490; MUID:89289996; PMID:2737288

A:Accession: S04490

A:Molecule type: protein

A:Residues: 94-128; 132-164 <MEN>

C:Superfamily: trypsin; trypsin homology

C:Keywords: glycoprotein; hydrolase; serine proteinase; zymogen

F:1-17/Domain: signal sequence #status predicted <SIG>

F:18-28/Domain: activation peptide #status predicted <ACT>

F:29-270/Product: (or 31-270) pancreatic elastase IIIB #status predicted <MAT>

F:29-263/Domain: trypsin homology <TRY>

F:73-123-217/Active site: His, Asp, Ser #status predicted

F:114/Binding site: carbonyl (Asn) (covalent) #status experimental

F:153/Binding site: carbonyl (Asn) (covalent) #status absent

Alignment Scores:

Pred. No.: 0.0458 Length: 270

Score: 114.50 Matches: 38

Percent Similarity: 37.04% Conservative: 22

Best Local Similarity: 23.46% Mismatches: 47

Query Match: 4.26% Indels: 55

DB: 2 Gaps: 3

US-10-020-441-1 (1-1477) x B29934 (1-270)

Qy 91 TGTACAGTTGACAGTGTTCAGAGACAGACAGTACAGTGTTCATTTGTTGCTCA 150

Db 58 CysGlyGlySerLeuIleAlaProAspTrpValIlnAlaGlyAlaCysIleSer 77

Qy 151 CCATTGCCAGTATCGGGTAAGAGATGACTGAAACATTTGTTCTCAATGTATGCA 210

Db 77 -----

Qy 211 TTGATTGCACAGACAGTGTTCGTTGCGTTCGTTGTGTGTGTGATGCAGTGGGTTT 270

Db 78 -----SerArgThr-----

Qy 271 GCATGATTTTAATGCTATATCTTACTTGTGCTTATTTTCAGTTTCTCAGCTGAGG 330

Db 81 -----TyrIlnValIlnLeuGlyIlnTyrAspArgAlaValIlns 93

Qy 331 AATGCGACCAACAGACGATCATCACCACCGTTCGAGTTAAGTGCACAGATAC 390

Db 94 GIuGlyProGIuGlnIlnValIlnProIlnsSerGIuAspIlnPheValIlnIleProIlnTrp 113

Qy 391 ATGCCCTTGTGTATGTCGGACAGACAGAGACCAATTCGACAGACACTGATGATTC 450

Db 114 AsnArgSerCysValAlaIlns-----GlyAsn 122

Qy 451 GATATTGCAATTTGATATGCTGCGTCAATGTCACCTTACAGATGATCAGATGATC 510

Db 123 AspIleAlaLeuIleLeuLeuSerIlnSerIlnSerIlnAlaIlnLeuGIuAspAlaValIlnLeuAla 142

Qy 511 AGTGTGCACACGATCGATATCCCGCACCTGGAACCTGTTGTTTCAATTTGTTGTTAT 570

Db 143 SerIlnProIlnAlaGIuAspIleLeuProAsnGIuTrpProCysTyrIlnThrGIuTrp 162

Qy 571 GGAAG 576

Db 163 GIuArg 164

RESULT 15

S54157

extensin-like protein - cowpea (fragment)

C:Species: Vigna unguiculata (cowpea)

C:Date: 08-Jul-1995 #sequence\_revision 03-Aug-1995 #text\_change 03-Aug-1995

C:Accession: S54157

R:Arzenjievic-Makelmovic, I.; Broughton, W.J.; Krause, A.

submitted to the EMBL Data Library, April 1995

A:Description: A class of root-hair specific extensins involved in rhizobium/legume int

A:Reference number: S54155

A:Accession: S54157

A:Molecule type: preliminary

A:Residues: 1-279 <AKS>

A:Cross-references: EMBL:X86028

Alignment Scores:

Pred. No.: 0.0628 Length: 279

Score: 113.00 Matches: 52

Percent Similarity: 35.56% Conservative: 28

Best Local Similarity: 23.11% Mismatches: 77

Query Match: 4.24% Indels: 68

DB: 2 Gaps: 9

US-10-020-441-1 (1-1477) x S54157 (1-279)

Qy 1120 ATCACTGGCACATATATCCACCCTTTACAC-----TACCACACACAAAT 1076

















hypothetical protein F59A6.3 - *Caenorhabditis elegans*  
 C/Species: *Caenorhabditis elegans*  
 C/Date: 20-Sep-1999 #sequence\_revision 20-Sep-1999 #text\_change 20-Sep-1999  
 C/Accession: T16509  
 R/Nhan, M.  
 submitted to the EMBL Data Library, December 1995  
 A/Description: The sequence of *C. elegans* cosmid F59A6.  
 A/Reference number: Z18526  
 A/Accession: T16509  
 A/Status: preliminary; translated from GB/EMBL/DBJ  
 A/Molecule type: DNA  
 A/Residues: 1-786 <NHA>  
 A/Cross-references: EMBL:U41994; NID:g1123047; PID:g1123052; PIDN:AAA83456.1; CESP:F59A6  
 A/Gene: CESP:F59A6.3  
 A/Introns: 106/1; 620/1; 634/3; 675/3; 775/1

Alignment Scores:  
 Pred. No.: 0.241 Length: 786  
 Score: 107.00 Matches: 80  
 Percent Similarity: 34.31% Conservative: 37  
 Best Local Similarity: 23.46% Mismatches: 104  
 Query Match: 4.01% Indels: 120  
 DB: Gaps: 18

US-10-020-441-1 (1-1477) x T16509 (1-786)

```

QY 1369 TGGACCTGGAAGGATGGAGAGAGCTCCACCACTGTCACCTGGAGCGGTAACGTGCC 1310
DB 69 TTPLeuLeuLysSerGluInu-----ThrThGluArgTrpAsnGly--LeuThr 84
QY 1309 GAATTCCTGACCTGCTTTCACACATATAGATTGCCATTGTCGTCATGCGCATTCAT 1250
DB 85 GlnPheLeuPheCysValLys-----GlyAlaCysSer----- 95
QY 1249 TATAGTCGTCGACCTGTGG-----TCCAAACAAACACCAACA----- 1208
DB 96 ---AspProSerProValTrpThrAnCySerThrGluThrThrSerThrIaSerThrGlu 114
QY 1207 ---ACAGCTACATGACACGATACGATACGATGAGAAAGAAACAAACACGAGACA 1151
DB 115 PheThrIle-SerSerThrSerLeuLysThrSerThrSerAspSerThrSerThrGlu-- 133
QY 1150 AAACAAACACACAGTGAATTCGAAATATCACTGGCACAATATCCACCCCTTACAC 1091
DB 134 -----ProArgIleSerThr-----Th 139
QY 1090 TACCCACACACAATATAGTCCACATCAATCAATGTCACATATTCACGATCAAGTCA 1031
DB 139 rThraSpThrLysAspThrThrThrGluAspProValSerSerThrAspGlnSerSerTh 159
QY 1030 CTCTCCA-----TCTCCACCACTCTCC 1010
DB 159 rSerProHisGluThrThrArgAspThrThrThrGluGlyThrThrSerGluAspSerTh 179
QY 1009 AAACACATACCTCAACACAC-----TTGACCCCTCTACTTCACTGTGA 971
DB 179 rSerThrTrpGlySerThrGluArgSerSerSerProLysProThrSerGluPheSerTh 199
QY 970 CACACACACCTTCATCGTTTGCATATGATATTTGATGAGTCTTCTCTCCACAGTGTC 911
DB 199 rSerThrGluSerAsp-----PheThrGluSerTh 209
QY 910 TCATATATCCATCATGATTCATCCATCCATCAATCTACTCAACTCACTCACTCAATC 851
DB 209 rArgSerSerGlu-ThrThrSerSerIleGluThrAsn-----SerSerThrSerProV 227
QY 850 ACCTGACCTCACTCAACACACACATGACGAGATGAATATCACTGTGCTAGTGAATCTG 791
DB 227 alSerThrThrProGluTrpAspSerThrSerSerGlyAsnSerGluThrThrGlu----- 245
QY 790 CACAAAGTCATCCAAACACAGTTGTTATTATCAAGGCTCTG----- 750

```

```

DB 246 -----SerAspGlyThrThrThrThrValPheThrThrThLysAspAspThrSerThrV 264
QY 749 -----ACCATACACAA 737
DB 264 alSerGlyAspSerAsnSerGlySerSerThrSerGluPheLysAsnThrThrThrT 284
QY 736 CCGGATCAGGT-----AGACACACACACACACAAACAG 704
DB 284 hGlyLProGlySerThrValSerGluProSerSerSerSerGluArgSerAspLeuAspSer 304
QY 703 ATACACACACACACAGATTAATCACTGACATCTGACGACGACTGATGATGCTTTA 644
DB 304 rSerValSerAspAspSerThrAsp-----S 313
QY 643 TTCACCAACAACTCACTTCTTCATATATTCAC---CCATTCCTTACGTACGATCAGCT 587
DB 313 rGlnAspArgThrThrGluIleGlyLeuGlnGlyProIleLeuSerAspAspSerAsnA 333
QY 586 CGTTATCATCCCTTCATACCAACATGAAACACCGCTTCAGTGCGGATATATCC 528
DB 333 sn-----ProAspProSerThrThrSerAlaLeuThrSerGlyThrSer 348

```

RESULT 30  
 T30283  
 polyketide synthase - *Streptomyces* sp. (strain MA6548)  
 C/Species: *Streptomyces* sp.  
 A/Variety: strain MA6548  
 C/Date: 22-Oct-1999 #sequence\_revision 22-Oct-1999 #text\_change 17-Nov-2000  
 C/Accession: T30283  
 R/Motamedi, H.; Cai, S.-J.; Shafiee, S.-J.; Elliston, K.O.  
 Eur. J. Biochem. 244, 74-80, 1997  
 A/Title: Structural organization of a multifunctional polyketide synthase involved in c  
 A/Reference number: Z20806; MUID:97217427; PMID:9063448  
 A/Accession: T30283  
 A/Status: preliminary; translated from GB/EMBL/DBJ  
 A/Molecule type: DNA  
 A/Residues: 1-6420 <NOT>  
 A/Cross-references: EMBL:Y10438; NID:e1014806; PID:e290681; PIDN:CAA71463.1  
 A/Genetics:  
 A/Note: fkbA  
 C/Superfamily: 3-oxoacyl-[acyl-carrier-protein] synthase I homology; acyl carrier prote  
 C/Keywords: carrier protein  
 F/51-433/Domain: 3-oxoacyl-[acyl-carrier-protein] synthase I homology <OAS1>  
 F/1930-2125/Domain: 3-oxoacyl-[acyl-carrier-protein] synthase I homology <OAS2>  
 F/3391-3462/Domain: acyl carrier protein homology <ACP1>  
 F/3505-3800/Domain: 3-oxoacyl-[acyl-carrier-protein] synthase I homology <OAS3>  
 F/3983-4254/Domain: [acyl-carrier-protein] S-malonyltransferase homology <AMT1>  
 F/5307-5378/Domain: acyl carrier protein homology <ACP2>  
 F/5431-5831/Domain: 3-oxoacyl-[acyl-carrier-protein] synthase I homology <OAS4>  
 F/5932-6206/Domain: [acyl-carrier-protein] S-malonyltransferase homology <AMT2>  
 F/6293-6364/Domain: acyl carrier protein homology <ACP3>

Alignment Scores:  
 Pred. No.: 0.362 Length: 6420  
 Score: 106.00 Matches: 62  
 Percent Similarity: 31.52% Conservative: 25  
 Best Local Similarity: 22.46% Mismatches: 104  
 Query Match: 3.98% Indels: 86  
 DB: Gaps: 10

US-10-020-441-1 (1-1477) x T30283 (1-6420)

```

QY 1230 GGTGCAACAAACACACACACAGTACATGACACGATGATGATGATGATGATGATGAT 1171
DB 2995 GlyAlaAspGluLeuProGluGlyTrpThrLeuIleThrAlaThrHisProAspAspPro 3014
QY 1170 GAACAAACACACAGGAACAAAC-----AAACACACACAGTGAATTCGAAATA 1120
DB 3015 AspaAspProThrAsnProHisAsnThrProThrArgThrHisThrGlnThrThrArgVal 3034
QY 1119 TCATCTGGCACATATATCAACCTTTACATAC---CACACACAAATATGATCCACATACAA 1063
DB 3035 LeuThrAlaLeuGlnHisHisIleuThrThrThrAsnHisThrLeuIleValHisThrThr 3054

```

```

QY 1062 TCA----- 1060
Db 3055 ThrAspProGlyAlaIaValThrGlyLeuThrArgThrAlaGlnAsnGlnHisPro 3074
QY 1059 ---CATGTGCACTAATTATTCGA----- 1042
Db 3075 GlyArgIleHisLeuIleGluThrHisIleHisProHisIleThrProLeuProLeuThrGlnLeu 3094
QY 1041 -----CGATCAAGTCACTGTCCATCTGACCACTGTCCAAACACATACT----- 999
Db 3095 ThrThrLeuHisGlnProHisIleLeuArgLeuThrAsnAsnThrLeuHisIleThrProHisIle 3114
QY 998 -----CAACACATTTGACCCCTACTTTCATGTACACACACACACTTCCATGCTTG 949
Db 3114 UThrProIleThrThrHisIleAsnThrThrThrThrProAsnThr----- 3130
QY 948 CATGATATTTTCATGAGTCTTCTCTCCAAAGTGCTCAATCATATCCCATCATACACA 889
Db 3131 -----ProProLeuAsnProAsnHisIleAlaIleLeuIleThr 3142
QY 888 TTCCATCCCATTCATCTACTCACTCACTCACTCACTCACTCACTCACTCACTCACTCA 829
Db 3142 rGlyGlySerGlyThrLeuAlaGlyIleLeuAlaArgHisLeuAsnHisProHisIleThr 3162
QY 828 C-----AATGACCGAGATGGAAATCACTGTGCTACTGAAT 793
Db 3162 rLeuLeuSerArgThrProProProProThrThrProGlyIleHisIleProCysAlaSer 3182
QY 792 GTCAACAAAGTCACCAACACAGTTGATTATTCAGAGCTGTGACCAAT-----AC 742
Db 3182 UThr-----AspProThrGlnIleThrGlnAlaLeuThrHisIleProGlnIle 3198
QY 741 AACAAACCGATCAGGTACAGACAGACAGACAAACAGTATACAGACAGACAGAAATAC 682
Db 3198 OleuThrGlyIlePheHisIleThrAlaAlaThrLeuAspArgAlaThrLeuThrAsnLeuThr 3218
QY 681 TGACTGACATCTGACTGACGACTG-----AG 656
Db 3218 rProGlnHisLeuThrThrThrThrLeuGlnProGlyAlaAspAlaAlaThrHisLeuHisIle 3238
QY 655 TCATGTCGTTTATTCACCAACACTCACTTCTTCATATTCACCC 610
Db 3238 HisThrGlnAsnGlnProLeuThrHisIlePheValLeuIleTyrSerSer 3253
RESULT 31
B27059
hypotheetical protein 2 - soybean (fragment)
C/Species: Glycine max (soybean)
C/Date: 19-Nov-1988 #sequence_revision 19-Nov-1988 #text_change 18-Jun-1993
C/Accession: B27059
R/From: H. U. J. Nap, J. P. J. Gloudemans, T. J. Stiekema, W. J. van Dam, H. J. Govers, F. J. Louw
Proc. Natl. Acad. Sci. U.S.A. 84, 4495-4499, 1987
A/Title: Characterization of cDNA for nodulin-75 of soybean: a gene product involved in
A/Reference number: A94164
A/Accession: B27059
A/Molecule type: mRNA
A/Residues: 1-203 <PRA>
A/Note: the authors translated the codon ATG for residue 177 as Leu
Alignment Scores:
Pred. No.: 0.291 Length: 203
Score: 105.50 Matches: 58
Percent Similarity: 41.12% Conservative: 30
Best Local Similarity: 27.10% Mismatches: 77
Query Match: 3.96% Indels: 49
DB: 2 Gaps: 10
US-10-020-441-1 (1-1477) x B27059 (1-203)
QY 1265 CATGTGCGCATTCATTATAGTGGCTGCAGCCTGTGTGTCGAAACAAACACACACAAAC 1206
Db 6 HisLeuSerIleTyrLeu-----LeuLeuMetArgAsnHisIleGlnAsn 20

```

```

QY 1205 AGTACATGACAAACATGATGATGAGAAAGAAACAAACAAACGGAACAAACA 1146
Db 21 -----ThrTyrLeuLeuMetArgAsn-ArgHis-----G1 30
QY 1145 AACACACAAAGTGAATTTGAAAATATGACGTGGACACATATTCACCCCTTTACATACC 1086
Db 30 nAsnThrHisLeuLeuMetArgAsn-----HisProMetArgIleHis 44
QY 1085 ACACAAATATG-----CCATATCATCATCATGTGACATATATATTCGAC 1041
Db 44 HisIleArgSerThrHisIleHisIleLeuMetArgSerHisIleGlnSerThrAsnHisIleLeuMetArg 64
QY 1040 GATCAAGTCACTGTCCATCTGACCACTGTCCAAACATCACTCAACACAA-----T 990
Db 64 ernHisIleGlnSerMetAsnHisIleLeuMetArgAsnHisIleGlnAsnThrHisIleLeu 84
QY 989 TGACCCCTTACTTCATGTACACACACACTTCATCGTTTGCGATGATATTTTCATGATC 930
Db 84 eArgSerHisIleGlnAsnThrHisIleLeuMetArgAsnHisIleGlnAsnThrHisIle 104
QY 929 TTCTGTCTCCAGTGTCAATGATCCATCATCATCATCATCATCATCATCATCATCAT 870
Db 104 lLeu-----MetArgSerHisIleGlnSerThrAsnHisIleLeuMetArgSer- 119
QY 869 CTTCAACTCACTTCATCATCATCATCATCATCATCATCATCATCATCATCATCATCAT 819
Db 120 HisGlnSerThrSerHisIleLeuMetArgSerHisIleGlnSerIleAsnHisIleLeuMetArg 139
QY 818 GATGAAATCACTGTGCTACTGAAATGTTCACAAATCAT-----CCA 777
Db 140 AsnHisIleGlnAsnThrAsnLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeu 159
QY 776 AACACAGTTGTTATATCAAGCT-----CTGACCAATATACAAACCGGATCAGGT 726
Db 160 AsnThrAsnLeuLeuMetArgSerHisIleGlnAsnThrAsnLeuProMetArgSerHis 179
QY 725 AGACACAGACACACAAACAGATACACAGAC 692
Db 180 HisGlnCys-ThrHisProLeuMetArgAsnHis 190
RESULT 32
S37595
mucin JUL10 - human (fragment)
C/Species: Homo sapiens (man)
C/Date: 06-Jan-1995 #sequence_revision 06-Jan-1995 #text_change 08-Oct-1999
C/Accession: S37595
R/Author: J.
submitted to the EMBL Data Library, September 1993
A/Reference number: S37593
A/Accession: S37595
A/Status: preliminary
A/Molecule type: mRNA
A/Residues: 1-330 <ADB>
A/Cross-references: EMBL:X74956; NID:G407052; PIDN:CAA52911.1; PID:G407053
Biochem. J. 293, 329-337, 1993
A/Title: Degenerate 87-base-pair tandem repeats create hydrophilic/hydrophobic alternat
A/Reference number: S35047; MUID:93345858; PMID:7916618
A/Accession: S35048
A/Status: preliminary
A/Molecule type: mRNA
A/Residues: 1-330 <DUP>
Alignment Scores:
Pred. No.: 0.375 Length: 330
Score: 104.50 Matches: 73
Percent Similarity: 37.63% Conservative: 32
Best Local Similarity: 26.16% Mismatches: 90
Query Match: 3.92% Indels: 85
DB: 2 Gaps: 14
US-10-020-441-1 (1-1477) x S37595 (1-330)

```

```

QY 1238 GACCGTGTGTGCAAAACAACCAACA-----CAACAGTACGTACA 1195
      |||||:::|||||
Db 44 AsprProValLeuHisProArgHisSerHisLeuAspProHisArgAlaGlnHisGlySerThr 63
QY 1194 ACAAGTACAGTACAGATGAGAAAGAAACAACAACGGAACGAACAAACAACAAG 1135
      |||||:::|||||
Db 64 ValThrValProThrGlySerGlnProThrAlaSerSerThrArgGly-----79
QY 1134 TGAATTGGAAATATACATGCGCAATATCCACCGCTTATACATGACCAACAATATA 1075
      ::|||:::|||||
Db 80 -----ThrLaGlyThrLeuLysValLeuThrSerAspHisHisThrHis 94
QY 1074 GTCCAC---ATACAATCAATGTCGATATATTCGACGATCAAGTCACTCTCATCT--- 1021
      |||||:::|||||
Db 95 SerHisGlnLeuGlnSerHisSerLeuGlnSerHisArgProCysAsnArgProSerSer 114
QY 1020 -----CAACACTCTCCAAACAATCTCAAAACAATTGACCCCTTCAATGTACACA 967
      |||||:::|||||
Db 115 ThrGlyLysHisSerHisHisThrHisSer-----TyrGln 126
QY 966 ACAACATTCATCGTTTGTGATGATATTTCAAGTCTCTCTCCGCAAGTGTCTGAA 907
      ::|||:::|||||
Db 127 ArgTyrSerHisProLeuPhe-----LeuProGlyThrAlaThrPheArg 141
QY 906 TCATCCCAATCACA-----TCACAATTCACATCCCAATCCATCTAC 868
      |||||:::|||||
Db 142 LeuSerGln-ThrThrThrProThrAlaThrMetSerThrAlaThrProSerSerThrArg 161
QY 867 TCAACTCACTCAATCACTGATCACTCACTCACTCAACAATGACCGAGATGAAATCA 808
      ::|||:::|||||
Db 161 GlnLys-ThrValHis-ThrSerThrValLeuThrThrThrAlaThrThrArgAlaThr 180
QY 807 CTCGTGTAAGTACGTGTCAAAAGTATCCAAACAAGTGTGTTATATCAAGGCTC---- 752
      |||||:::|||||
Db 180 HisAlaLeuThrProProProProProGlnGlnGlnLeuThrLeuProLysCysArgL 200
QY 751 -----TGACAATACAAACAACG-----G 733
      |||||:::|||||
Db 200 LeuProGlnProArgLeuHisSerTyrProLeuLeuGlnProArgArgLysThrHisAla 220
QY 732 ATCAGGTACAGACAGACAGACAAACAGATACAGACAGACAGACAGATATGACTGACA 673
      |||||:::|||||
Db 220 SerSerValArgGlnHisAsnHis-----HisThrHisAsnGlnArgLeuHisGly- 236
QY 672 TGTGACTGACGAGTACGATGTCGTTTATTCACCAACATCACTTCTTCATTTGCTCC 613
      ::|||:::|||||
Db 237 -----AspPro-LeuLeuHisProGlyThrThrHis-----ThrAla 248
QY 612 ACCATTCTTACGTCGACGATCAG-----589
      |||||:::|||||
Db 249 ThrValLeuThrThrThrThrThrThrValProLeuValLeuThrGlnHisProProLeu 268
QY 588 -----GTGCTTATCATCCCTTCATACCAACAATGAAAAACAC 550
      |||||:::|||||
Db 269 AlaHisArgProValValLeuProProSerLeuThrThrThrAlaThrThr 285

RESULT 33
A48292
C1Species: Canis lupus familiaris (dog)
C1Date: 27-Jun-1994 #sequence_revision 27-Jun-1994 #text_change 05-Nov-1999
C1Accession: A48292
R1Verma, M.; Davidson, E. A.
Proc. Natl. Acad. Sci. U.S.A. 90, 7144-7148, 1993
A1Title: Molecular cloning and sequencing of a canine tracheobronchial mucin cDNA contai
A1Accession: A48292
A1Status: preliminary
A1Molecule type: mRNA
A1Residues: 1-1118 <VER>
A1Cross-references: GB:X69164; NID:9402557; PIDN:CAA48914.1; PID:9402558
*A1Note: authors translated the codon CAA for residue 52 as Gly, GAT for residue 94 as As

```

```

C1Comment: It is uncertain whether Met-1 or Met-52 is the initiator.
C1Keywords: glycoprotein

Alignment Scores:
Pred. No.: 0.42
Score: 104.50
Percent Similarity: 34.22%
Best Local Similarity: 24.89%
Query Match: 3.92%
DB: 2 Gaps: 6

US-10-020-441-1 (1-1477) x A48292 (1-1118)

QY 1219 ACAACCAACAACAGTACAGTACAAACGATACAGTACAGTACAGTACAGTACAGTAC 1160
      |||||:::|||||
Db 370 ThrThrSerThrLysVal-LeuValThrProTyrThrThrThrThrGlyValThrProTh 389
QY 1159 AACGGAACAAACAAACAAACACACAGTGAATTTCCAAAATATCACTGGCAATATCCAC 1100
      |||||:::|||||
Db 389 rThrHisArgHisThrArgHisGln-----ThrThrThrThrHisHis 403
QY 1099 CCGTTACATCCCAACAACAATATGTCACATACATACATGATGCACTATATTCGACG 1040
      |||||:::|||||
Db 403 rThrThrThrThrValThrAlaIleAla-----412
QY 1039 ATCAAGTCACTCTCCATCACTCACTCACTCAACAACATATCAACAACATGACCCCTA 980
      ::|||:::|||||
Db 413 -----ThrProThrGlyThrArgProThrThrThrThrThrThrHisThrThrThr- 429
QY 979 CTTCATGTACACAAACAATTCATCGTTTGTGATGATATTTCAAGTCTTCTCTCTC 920
      ::|||:::|||||
Db 429 alThrGlnThrThrHisArgGlnThrThr-----ArgProG 441
QY 919 CAAGGCTGTAATATCCCAATCAATCAATTCACATCCCAATCCATCACTCACTCACTCA 860
      |||||:::|||||
Db 441 LnPThrThrProSerProThrThrThrThrValAspProThrGlnHisProSerProP 461
QY 859 CCTCAATCACTCACTCACTCACTCACTCAACAACATGACCGAGATGAAATCACTGTGTA 800
      ::|||:::|||||
Db 461 rProLeuGlyAspProAsnProAsnThrHisArgAspThrAspProAsnHisAspThr 481
QY 799 CTGAATCG-----TCACAAGTATCCAAACA 773
      |||||:::|||||
Db 481 hrThrThrThrThrValThrProThrProThrProThrArgThrArgProGlnProArg 501
QY 772 CAGTTGTATATTCAGAGGCTGACCAATATCAACAACCGGATCAGTACAGACAGACA 713
      ::|||:::|||||
Db 501 yTyrThrThrThrThrThrMetThrProThrProThrProThrProThrHisThrAsp 521
QY 712 CACAAACAGAT-----ACAACAGACAGACAGATATACATGAA-- 679
      |||||:::|||||
Db 521 iHisHisAspProLeuProProProProProGlnValArgHisHisArgThrProArgH 541
QY 678 -----CTGACATCTGACGATGACGATGATGCTTTATTCACCAACAATCTC 630
      |||||:::|||||
Db 541 iCysHisGlnLeuThrSerGlySerProLeuThrHisGlyrProAsnThrProThrTha 561
QY 629 ACTTCTTCAA 619
      ::|||:::|||||
Db 561 enLeuLeuGln 564

RESULT 34
A43932
C1Species: Homo sapiens (man)
C1Date: 10-Mar-1993 #sequence_revision 12-Apr-1996 #text_change 05-Nov-1999
C1Accession: A43963; A45106; B45106; B35532; A61257; P00328; P00329
R1Gum Jr., J.R.; Hicks, J.W.; Toribara, N.W.; Siddiki, B.; Kim, Y.S.
J. Biol. Chem. 269, 2440-2446, 1994
A1Title: Molecular cloning of human intestinal mucin (MUC2) cDNA. Identification of the
A1Accession: A43963; MUID:94132002; PMID:8300571

```

A.Molecule type: mRNA  
A.Residues: 1-639 <GU1>  
A.Cross-references: GB:I21998  
R.Gum Jr., J.R.; Hicks, J.W.; Toribara, N.W.; Rothe, E.M.; Lagace, R.E.; Kim, Y.S.  
J. Biol. Chem. 267, 21375-21383, 1992  
A.Title: The human MUC2 intestinal mucin has cysteine-rich subdomains located both upstream and downstream of the mucin core  
A.Reference number: A45106; MUID:9916075; PMID:1400449  
A.Accession: A45106  
A.Status: not compared with conceptual translation  
A.Molecule type: mRNA  
A.Residues: 626-1895 <GU2>  
A.Cross-references: GB:I99431; NID:G186395; PIDN:AA59163.1; PID:G186396  
A.Note: sequence extracted from NCBI backbone (NCBIP:116706)  
A.Accession: B45106  
A.Status: not compared with conceptual translation  
A.Molecule type: mRNA  
A.Residues: 2037-3020 <GU3>  
A.Cross-references: GB:I99432; NID:G186397; PIDN:AA59164.1; PID:G186398  
A.Experimental source: colon  
A.Note: sequence extracted from NCBI backbone (NCBIP:116698)  
R.Toribara, N.W.; Gum Jr., J.R.; Cullane, P.J.; Lagace, R.E.; Hicks, J.W.; Petersen, G.M.  
J. Clin. Invest. 88, 1005-1013, 1991  
A.Title: MUC-2 human small intestinal mucin gene structure. Repeated arrays and polymorphism  
A.Reference number: A43932; MUID:91358717; PMID:1885763  
A.Accession: A43932  
A.Molecule type: DNA  
A.Residues: 1343-1350 'U', 1352-1411 'S', 1413-1448 'P', 1450-1503 'T', 1505-1915 <TOR>  
A.Cross-references: GB:I74027; NID:G188863; PIDN:AA59875.1; PID:G188864  
A.Note: sequence inconsistent with the nucleotide translation  
A.Note: sequence extracted from NCBI backbone (NCBIN:55749, NCBIP:55750)  
R.Gum, J.R.; Byrd, J.C.; Hicks, J.W.; Toribara, N.W.; Lampert, D.T.A.; Kilm, Y.S.  
J. Biol. Chem. 264, 6480-6487, 1989  
A.Title: Molecular cloning of human intestinal mucin cDNAs. Sequence analysis and evidence for a family of genes  
A.Reference number: A33532; MUID:89197956; PMID:2703501  
A.Accession: B33532  
A.Molecule type: mRNA  
A.Residues: 1916-2193 <GU4>  
A.Cross-references: GB:I22405; NID:G188873; PIDN:AA36334.1; PID:G188874  
A.Experimental source: intestine  
R.Juan, B.H.; Gallup, M.W.; Yan, P.S.; Gum, J.R.; Kilm, Y.S.; Basbaum, C.B.  
J. Clin. Invest. 87, 77-82, 1991  
A.Title: Human bronchus and intestine express the same mucin gene.  
A.Reference number: A61257; MUID:91086481; PMID:1985113  
A.Accession: A61257  
A.Status: not compared with conceptual translation  
A.Molecule type: mRNA  
A.Residues: 'T', 1925-1948 'TTS', 1952-1954 <JUN>  
A.Experimental source: bronchus  
R.Xu, G.; Huan, L.; Khatri, I.; Sajjan, U.S.; McCoil, D.; Wang, D.; Jones, C.; Forstner, B.  
Biochem. Biophys. Res. Commun. 183, 821-828, 1992  
A.Title: Human intestinal mucin-like protein (MLP) is homologous with rat MLP in the C-terminal region  
A.Reference number: PQ0328; MUID:92198477; PMID:1550588  
A.Accession: PQ0328  
A.Molecule type: mRNA  
A.Residues: 2328-2468 <XUG>  
A.Cross-references: GB:M86523  
A.Experimental source: small intestine  
A.Accession: PQ0329  
A.Molecule type: protein  
A.Residues: 2328-2342 'K', 2344-2354 <XUG1>  
C.Genetic8  
A.Gene: GDB:MUC2  
A.Cross-references: GDB:120203; OMIM:158370  
A.Map position: 11p15.5-1p15.5  
C.Superfamily: von Willebrand factor; von Willebrand factor type A repeat homology; von C-keywords: glycoprotein; intestine; tandem repeat  
F:2766-2834/Domain: von Willebrand factor type C repeat homology <VWC>  
  
Alignment Scores:  
Pred. No.: 0.461 Length: 3020  
Score: 104.50 Matches: 84  
Percent Similarity: 30.33% Conservative: 27  
Best Local Similarity: 22.95% Mismatches: 169

Query Match:	3..92%	Indels:	86
DB:	2	Gaps:	12
US-10-020-441-1 (1-1477) x A43932 (1-3020)			
QY 1392 CCATGTGATACGACACCGGATCTAGCACTTGAAGGAGTGGAGAGAGAGTCCACACTG 1333			
Db 1557 ProProthrThrProSerProProthThrThrThrThrThrThrThrThrThrThr 1576			
QY 1332 TCACCTGAGCGGGTAACTGTCCGAAATTCGACCTTGCTTTACACATATAGATTGCCA 1273			
Db 1577 ThrProSerPro-----ProthThrThrThrProSerPro 1589			
QY 1272 TTGGTCGATGTGGGCACTTCATTATAGTGGCGCACTGTGGTGAACAAACAAACCA 1211			
Db 1590 ThrThrThrThr-----ThrThrProProPro 1599			
QY 1212 CAACACAGTACAGTACACACGTAAGTACAGATGAGAGAGAGAGAGAGAGAGAGAG 1153			
Db 1599 ThrThrThrProSerProProthThrThrThrThrThrThrThrThrThrThrPro 1618			
QY 1152 CAACACAGAGTACAGATGATTTGG---AAATATCACTGGGACATATACACCTT 1099			
Db 1619 SerProProthThrThrThrProthThrThrThrThrThrThrThrThrThrProth 1638			
QY 1095 TACACTCCGACCAACAAATAGCCACATACATACATACATGTCACATATATTCAGATCA 1036			
Db 1638 ThrThrThrProSerProProthThrThr----- 1647			
QY 1035 AGTCACTCTCCATCTCCACCACTCTCCAAACACATACATCAACACATGACCCCTACTTC 976			
Db 1648 -----ThrThrThrProProthThrThrThrThrThrThrThrThrThrThrThr 1662			
QY 975 ATGACACACACACACTTCCATCGTTGGTCATAGATATTTATAGAGCTTCTCTCCACAG 916			
Db 1662 rThThrProSerProProthThrThrThrThrThrThrThrThrThrThrThrThrThrProse 1682			
QY 915 TGGTCAATCATCCATCCATCATCATCATCATCCATCCATCATCTCAACTCACTCC 856			
Db 1682 rSer-----ProthThrThrThrThrThrThrThrThrThrThrThrThrThrThrThrPr 1699			
QY 855 ACATCACTGACCTCACTCACTCCACCAACAAATGACCGAGATGAGAAATCACTGTGCTACTGA 796			
Db 1699 oSerProthThrThrThrProSerSerProthThrThrThrThrThrThrThrThrThrThr 1719			
QY 795 ACTGTCAAAAGTCATCCAAACACAGTGTATTATCAAGGCTGTGACCAATACAAACAC 736			
Db 1719 rThThrProSerProProthThrThrThrThr-----ThrProSerProthThrThrThr 1736			
QY 735 CGATTCAGGTAGACAGACACACACAAACAGATACACAGACACAGAAATACGACTG 676			
Db 1736 rProSerProProthThrThrThrThrThrThrThrThrThrThrThrThrThrThrThrThr 1750			
QY 675 ACATCTGACTGACGACTGACTGCTGTTATTACCAACAACTCACTTCTTCAATAT 616			
Db 1751 -----ThrThrSerSerProleuthThrThrThrProle 1761			
QY 615 TCACCACTTCTTAGTGAAGGATACAGGTCGTATTATCCCTTCCATATACCAACAAATGAA 556			
Db 1761 uProProSerThrThrThrThrThrThrThrThrThrThrThrThrThrThrThrThrThrThr 1779			
QY 555 AACACCAAGT----- 546			
Db 1780 -ThrProCysValProleuCysAsnTrpThrGlyTrpLeuAspSerGlyLysProAsnPh 1799			
QY 545 -----CAAGTGGCGGG-----ATATCCGATGGCTGTGGCAAGCTGATCATCTT 502			
Db 1799 eHisLysProGlyGlyAspTrhGluLeuIleGlyAspValCysGlyProGlyTrpAlaAl 1819			
QY 501 GATTTCACCTCGTAAGTGTACCAATTGAGCCAGCAATTACATGCAATATGGAATCCACT 442			
Db 1819 aAsnIleSerCysArgValAlaThrMetCysProAspArgProIleGlyGluLeuGlyGlnTh 1839			







```

Db      163 Gly-----:::  :::  |||  |||||  :::|  |||
      AlaproValasphrThrThrIleArgasnValTyrProLeuPro 178

Qy      1083 AACAAATAGTCACATACATACATGACCTATATTCAGCATCAAGTACACTCCA 1024
      |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db      179 ValThrThrThrGluThrGlnAlaProProHisThrThrLys--ProTyrGlnTyr 197

Qy      1023 TCTCAGCACTCTCCAAACACATCTCAAAACATTCAGCCCTTCTGATGACAAACA 964
      :::  |||||  |||  |||  |||  |||||  |||||  |||||  |||||
Db      198 HisTyrThrSerProThrThr--ThrThrThrLysArgProProThrThrThrThrGln 217

Qy      963 CACTTCACATCGTTGGATAGATATTTCATGAGTCTTCTCTCTCCAAAGTGTCAATCA 904

Db      217 uasPro-----G1 220

Qy      903 TCCCATCATCATCATTCATC-----CCATC 877

Db      220 nProheGluPheHisGluLeuGlnGlnAlaGluasnProIleProLeuasnProIly 240

Qy      876 CAATCTACCTCAACTCACTCATCATCTCAGTCACTCAGCTCAACACATGACCGAGA 817
      |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db      240 sasnSerIeuGlnGlnThrSerGluGlnVal--AspGluGlnProPheasnPhelysG1 259

Qy      816 TGGAAATCACTCTGCTACTGAACTGCA--CAAAGTCATCCAAACACAGTTGTTATTAT 760
      :  :::  |||||  |||  |||  |||  |||  |||  |||  |||
Db      259 uProThrAsnSerAlaSerAspGlySerGlnGlnSerValGluSerThrThrThrTh 279

Qy      759 CAAGCTCTGACC-----AATACAAACACCGGATCAGGTAGACAGACAGACACA 709
      :::  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db      279 rGluSerProThrThrGluAlaSerThrThrThrThrGluAlaProThrThrThrTh 299

Qy      708 AACAGATACACAGACACAGATATGACTGATGATGATGATGATGATGATGATGATG 649
      :::  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db      299 rGluGlnValThrThrThrThrThrThrThrThrThrThrThrThrThrThrThrThr 318

Qy      648 GTTATATCCAAACATCTTCTTCATATTCACCATTTCTTACGTCGATCAG 589
      :::  |||||  |||  |||  |||  |||  |||  |||  |||  |||
Db      319 ThrThrThrGluThrThrThrThrThrThrThrThrThrThrThrThrThrThrThr 333

Qy      588 GTTCGTTATCATCCCTTCCTTCAATACCAACATGAAAACACAGATGCC 544
      :::  |||||  |||  |||  |||  |||  |||  |||  |||  |||
Db      334 GluArgMetSerProSerThrThrThrThrThrThrThrThrThrThrThrThr 348
  
```

```

RESULT 39
S50392
hypothetical protein YMR136w - yeast (Saccharomyces cerevisiae)
N:Alternate names: hypothetical protein YMR375.05
C:Species: Saccharomyces cerevisiae
C>Date: 13-Jan-1995 #sequence_revision 24-Feb-1995 #text_change 19-Apr-2002
C:Accession: S50392
R:Badcock, K.; Churcher, C.
submitted to the EMBL Data Library, December 1994
A:Reference number: S50388
A:Accession: S50392
A:Molecule type: DNA
A:Residues: 1-560 <BAD>
A:Cross-references: EMBL:Z47071; NID:g606429; PIDN:CAA87350.1; PID:g606434; GSPDB:GN0001
C:Genetics:
A:Gene: SGD:GAT2; MIPS:YMR136w
A:Cross-references: SGD:S0004744
A:Map position: 13R
  
```

```

Alignment Scores:
Pred. No.: 0.485 Length: 560
Score: 103.50 Matches: 72
Percent Similarity: 39.86% Conservative: 44
Best Local Similarity: 24.74% Mismatches: 81
Query Match: 3.88% Indels: 94
DB: 2 Gaps: 17
  
```

US-10-020-441-1 (1-1477) x S50392 (1-560)

\* Qy 1347 AGAGGTCCACCACTGTCACTGGAGCGGGTAACTGCGAATAATTCGACCTGCTTTCACA 1288

```

Db      2 GlnAlaProAsnIleTyrProPheSerGlnThrGlnProGlnAlaLeuProGlyPheThr 21
      :::  |||  :::  |||  :::  |||||  |||  |||||
Qy      1287 CATATAGGATTTGCCATTGCGCATGTCGATTCATTAATAGTCGACCGCTGGT 1228
      |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db      22 Tyr--GlyProProGlnLeuValPheAspHisSerAlaPro-ArgValaProLeu-- 39

Qy      1227 GCMAACAAACCAACCAACCAACAGTACAGTACCAAC--GTACAGTACAGAT 1177
      :::  |||||  |||  |||  |||  |||  |||  |||  |||  |||
Db      40 -----His-SerThrValThrIleAsnSerProLeuProLeuGlnHis-- 53

Qy      1176 GAGAAAGACAAACACACAGGAAACAAACAAACACAAAGTAAATTCCAAAAATATCA 1117
      :::  |||||  |||  |||  |||  |||  |||  |||  |||  |||
Db      54 -----TyrAsnGly-----ProAsnAlaHisIleAsnSerAlaAsnAsn 67

Qy      1116 ACTGGACATATATGACCTTTACCTTACCTACGACCAACATAGTGCATCAATACAT 1057
      |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db      67 snTyr-----AlaTyrTyrTyrHisIleProAsnAsnAsnAsnAsnHisSer- 84

Qy      1056 GTGCATATATTCAGATCAAGTCACTCTCCATCTCACTCACTCAACATATCTCA 997
      :::  |||||  |||  |||  |||  |||  |||  |||  |||  |||
Db      85 -----AsnAsnThrIleLysAsnAsnAsnIleAsnSerValLeuProAlaValAsnI 102

Qy      996 AACACATTGACCCCTACTTATGATGTACAAACACACTTCATCGTTGATAGATATTTC 937
      |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db      102 leGlnIleSerAsnAsnSerHisTyrArgAsnThr-----H 114

Qy      936 ATGAG-----TCTTCTCTCTCCAGAGTGTCTCAATCCATCCATCATCATTCAT 883
      |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db      114 isGlnIleProSerAlaProGlnArgLeuValSerIleIleProAsnProHisMetPro- 133

Qy      882 CCCATCAATCTTACCTCACTCACTCACTCACTCACTCACTCACTCACTCACTCACTCA 823
      |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db      134 -----ProAsnIleSerHisPheGlnLeuAsnIleHisP 146

Qy      822 CCGAGATGAATATCACTGTCTACTGAACTGTCAGAAAGTATCCAAACACAGTTGTTAT 763
      :::  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db      146 roGlnMet-----HisAlaProValAlaThrAspIle-His 157

Qy      762 TATCAA-----GGCTGACCAATATAC 739
      |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db      158 PheGlnGlnValProValTyrAsnLysThrAsnAsnGlyIleGlyThrAspAsnIleAsn 177

Qy      738 AACCGATCAGGTAGACAGACAGACACAAACAGATACACAGACACAGATATCTGA 679
      |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db      178 Asn-----AspLysProValAsn-----SerAsnGlnAsnGlnVal 189

Qy      678 CTGACATCT--GACTGACGACTGATGATGTC-----GTTATTTCCAAACAACT 631
      |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db      190 LeuAspAsnIleAspGluArgSerCysHisGluIleAsnArgValaValSerPheSerLys 209

Qy      630 CACTTCTTCAATATATTCACCATTTCTTACGTCAGGATCAGGTGTTATATCTTCCTTC 571
      |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db      210 HisPhe----- 211

Qy      570 ATAACCAACATGAAGAAACACAGTTC 544
      |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db      212 ---GluAsnAsnGluLeuThrThrThr 219
  
```

```

RESULT 40
C56653
membrane glycoprotein precursor - human herpesvirus 6 (strain UI102)
C:Species: human herpesvirus 6
C>Date: 05-Jan-1996 #sequence_revision 05-Jan-1996 #text_change 21-Jul-2000
C:Accession: C56653
R:Compels, U.A.; Cars, A.L.; Sun, N.; Arrand, J.R.
DNA Seq. 3, 25-39, 1992
A:Title: Infectivity determinants encoded in a conserved gene block of human herpesviri
A:Reference number: A56653; MUID:93091236; PMID:1333836
A:Accession: C56653
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-651 <GDM>
  
```



**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.4 p5.4578  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 9, 2003, 12:58:12 ; Search time 5.90164 Seconds

(without alignments)  
165.747 Million cell updates/sec

Title: US-10-020-441-3

Perfect score: 90

Sequence: 1 VGYGRDNDNDPDSRKN 16

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 248812 seqs, 61136040 residues

Total number of hits satisfying chosen parameters: 248812

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 10%

Listing first 45 summaries

Database : Published Applications\_AA:\*  
1: /cgn2\_6/ptodata/2/pubppaa/US08\_NEW\_PUB pep:\*  
2: /cgn2\_6/ptodata/2/pubppaa/PCT\_NEW\_PUB pep:\*  
3: /cgn2\_6/ptodata/2/pubppaa/US06\_NEW\_PUB pep:\*  
4: /cgn2\_6/ptodata/2/pubppaa/US06\_PUBCOMB pep:\*  
5: /cgn2\_6/ptodata/2/pubppaa/US07\_NEW\_PUB pep:\*  
6: /cgn2\_6/ptodata/2/pubppaa/US07\_PUBCOMB pep:\*  
7: /cgn2\_6/ptodata/2/pubppaa/PCTUS\_PUBCOMB pep:\*  
8: /cgn2\_6/ptodata/2/pubppaa/US08\_PUBCOMB pep:\*  
9: /cgn2\_6/ptodata/2/pubppaa/US09\_NEW\_PUB pep:\*  
10: /cgn2\_6/ptodata/2/pubppaa/US09\_PUBCOMB pep:\*  
11: /cgn2\_6/ptodata/2/pubppaa/US10\_NEW\_PUB pep:\*  
12: /cgn2\_6/ptodata/2/pubppaa/US10\_PUBCOMB pep:\*  
13: /cgn2\_6/ptodata/2/pubppaa/US60\_NEW\_PUB pep:\*  
14: /cgn2\_6/ptodata/2/pubppaa/US60\_PUBCOMB pep:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	90	100.0	16	9 US-10-020-441-3	Sequence 3, Appli
2	90	100.0	106	9 US-10-020-441-2	Sequence 2, Appli
3	44	48.9	614	9 US-10-151-557-2	Sequence 2, Appli
4	42	46.7	500	10 US-09-764-864-1362	Sequence 1362, Ap
5	42	46.7	1005	10 US-09-925-301-1335	Sequence 1335, Ap
6	41	45.6	229	9 US-09-764-868-847	Sequence 847, App
7	40.5	45.0	229	9 US-09-894-912A-25	Sequence 25, Appl
8	40.5	45.0	265	9 US-10-125-852-24	Sequence 24, Appl
9	40.5	45.0	829	10 US-09-815-242-11216	Sequence 11216, A
10	40	44.4	228	9 US-09-866-050A-684	Sequence 684, App
11	39	43.3	104	9 US-09-870-759-90	Sequence 90, Appl
12	38	42.2	17	9 US-10-083-590-12	Sequence 12, Appl
13	38	42.2	17	9 US-09-925-287-1	Sequence 1, Appli
14	38	42.2	35	9 US-10-083-815-68	Sequence 68, Appl
15	38	42.2	107	9 US-10-083-815-71	Sequence 71, Appl
16	38	42.2	176	10 US-09-864-761-35255	Sequence 35255, A
17	38	42.2	193	10 US-09-925-301-983	Sequence 983, App
18	38	42.2	445	9 US-09-286-488-38	Sequence 38, Appl
19	38	42.2	445	10 US-09-737-178-38	Sequence 38, Appl

20	38	42.2	482	9 US-09-738-626-4373	Sequence 4373, Ap
21	38	42.2	523	9 US-10-138-838-95	Sequence 95, Appl
22	38	42.2	523	9 US-10-138-838-110	Sequence 110, Appl
23	38	42.2	523	9 US-10-139-031-95	Sequence 95, Appl
24	38	42.2	523	9 US-10-139-031-110	Sequence 110, Appl
25	38	42.2	666	10 US-09-737-178-85	Sequence 85, Appl
26	38	42.2	677	10 US-09-737-178-144	Sequence 144, Appl
27	38	42.2	1132	10 US-09-737-178-87	Sequence 87, Appl
28	37	41.1	31	10 US-09-864-761-36056	Sequence 36056, A
29	37	41.1	42	10 US-09-864-761-35018	Sequence 35018, A
30	37	41.1	104	9 US-09-983-802-172	Sequence 172, App
31	37	41.1	333	10 US-09-815-242-10900	Sequence 10900, A
32	37	41.1	774	10 US-09-815-242-12046	Sequence 12046, A
33	37	41.1	829	10 US-09-801-368-324	Sequence 324, App
34	37	41.1	1004	9 US-09-738-626-5676	Sequence 5676, App
35	37	41.1	1285	10 US-09-982-091A-2	Sequence 2, Appli
36	37	41.1	5179	9 US-10-025-380-1068	Sequence 1068, Ap
37	37	41.1	5179	10 US-09-922-217-1068	Sequence 1068, Ap
38	37	41.1	5179	10 US-09-833-263-1068	Sequence 1068, Ap
39	36.5	40.6	69	10 US-09-864-761-47955	Sequence 47955, A
40	36.5	40.6	447	10 US-09-741-669-405	Sequence 405, App
41	36	40.0	30	10 US-09-864-761-47769	Sequence 47769, A
42	36	40.0	62	10 US-09-864-761-36587	Sequence 36587, A
43	36	40.0	84	10 US-09-864-761-33814	Sequence 33814, A
44	36	40.0	85	10 US-09-016-869B-34	Sequence 34, Appl
45	36	40.0	106	9 US-09-764-868-1168	Sequence 1168, Ap

## ALIGNMENTS

```
RESULT 1
US-10-020-441-3
; Sequence 3, Application US/10020441
; Publication No. US2002018224A1
; GENERAL INFORMATION:
; APPLICANT: Doenhoff, Michael
; TITLE OF INVENTION: SCHISTOSOMIASIS VACCINE
; FILE REFERENCE: Doenhoff
; CURRENT APPLICATION NUMBER: US/10/020,441
; PRIOR FILING DATE: 2001-12-18
; PRIOR APPLICATION NUMBER: US/09/413,810
; NUMBER OF SEQ ID NOS: 3
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 3
; LENGTH: 16
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: protein
US-10-020-441-3

Query Match          100.0%; Score 90; DB 9; Length 16;
Best Local Similarity 100.0%; Pred. No. 2e-08;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 VGYGRDNDNDPDSRKN 16
|||
Db      1 VGYGRDNDNDPDSRKN 16

RESULT 2
US-10-020-441-2
; Sequence 2, Application US/10020441
; Publication No. US2002018224A1
; GENERAL INFORMATION:
; APPLICANT: Doenhoff, Michael
; TITLE OF INVENTION: SCHISTOSOMIASIS VACCINE
; FILE REFERENCE: Doenhoff
; CURRENT APPLICATION NUMBER: US/10/020,441
```

```
; CURRENT FILING DATE: 2001-12-18
; PRIOR APPLICATION NUMBER: US/09/413,810
; PRIOR FILING DATE: PRIOR FILING DATE: 1999-10-07
; NUMBER OF SEQ ID NOS: 3
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2
; LENGTH: 106
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: protein
US-10-020-441-2
```

```
Query Match          100.0%; Score 90; DB 9; Length 106;
Best Local Similarity 100.0%; Pred. No. 1.6e-07;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 1 VVGGRDNDRDPSRK 16
    |||||
Db 85 VVGGRDNDRDPSRK 100
```

```
RESULT 3
US-10-151-557-2
; Sequence 2, Application US/10151557
; Publication No. US20020183506A1
; GENERAL INFORMATION:
; APPLICANT: No. US20020183506A1:ozymes A/S
; APPLICANT: Danielson, Steffen
; APPLICANT: Schneider, Falle
; TITLE OF INVENTION: Nucleic acids encoding polypeptides having
; FILE REFERENCE: 10173.200-US
; CURRENT APPLICATION NUMBER: US/10/151,557
; CURRENT FILING DATE: 2002-05-18
; PRIOR APPLICATION NUMBER: US/09/832,614A
; PRIOR FILING DATE: 2001-04-11
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 2
; LENGTH: 614
; TYPE: PRT
; ORGANISM: Geniculosporium sp.
US-10-151-557-2
```

```
Query Match          48.9%; Score 44; DB 9; Length 614;
Best Local Similarity 80.0%; Pred. No. 26;
Matches 8; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
```

```
QY 5 RDDNDRDPSR 14
    |||||
Db 366 RDDNFRDPR 375
```

```
RESULT 4
US-09-764-864-1362
; Sequence 1362, Application US/09764864
; Patent No. US20020132753A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PT23
; CURRENT APPLICATION NUMBER: US/09/764,864
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - consult PALM or file wrapper
; NUMBER OF SEQ ID NOS: 1792
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1362
; LENGTH: 500
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
```

```
; LOCATION: (154)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (167)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (198)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (203)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-09-764-864-1362
```

```
Query Match          46.7%; Score 42; DB 10; Length 500;
Best Local Similarity 66.7%; Pred. No. 43;
Matches 8; Conservative 1; Mismatches 3; Indels 0; Gaps 0;
```

```
QY 4 GRDNDRDPSRK 15
    |||||
Db 188 GRGTHDRDPSXK 199
```

```
RESULT 5
US-09-925-301-1335
; Sequence 1335, Application US/09925301
; Patent No. US20020052308A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
; FILE REFERENCE: PA106
; CURRENT APPLICATION NUMBER: US/09/925,301
; CURRENT FILING DATE: 2001-08-10
; PRIOR APPLICATION NUMBER: PCT/US00/05882
; PRIOR FILING DATE: 2000-03-08
; PRIOR APPLICATION NUMBER: 60/124,270
; PRIOR FILING DATE: 1999-03-12
; NUMBER OF SEQ ID NOS: 1694
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1335
; LENGTH: 1005
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-925-301-1335
```

```
Query Match          46.7%; Score 42; DB 10; Length 1005;
Best Local Similarity 54.5%; Pred. No. 91;
Matches 6; Conservative 4; Mismatches 1; Indels 0; Gaps 0;
```

```
QY 5 RDDNDRDPSRK 15
    |||||
Db 839 REENDKDPERE 849
```

```
RESULT 6
US-09-764-868-847
; Sequence 847, Application US/09764868
; Patent No. US20020168711A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PT232
; CURRENT APPLICATION NUMBER: US/09/764,868
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - refer to PALM or file wrapper
; NUMBER OF SEQ ID NOS: 1510
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 847
; LENGTH: 229
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (119)
```

OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids  
US-09-764-868-847

Query Match 45.6%; Score 41; DB 9; Length 229;  
Best Local Similarity 50.0%; Pred. No. 27;  
Matches 7; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

QY 3 YGRDNDNRPSRKN 16  
: : : : :  
Db 90 FGKXKXDRSSVRN 103

RESULT 7  
US-09-894-912A-25  
Sequence 25, Application US/09894912A  
Publication No. US20030044792A1  
GENERAL INFORMATION:  
APPLICANT: Tang et al.  
TITLE OF INVENTION: METHODS AND MATERIALS RELATING TO NOVEL STEM CELL GROWTH  
FILE REFERENCE: 28110/37260A  
CURRENT APPLICATION NUMBER: US/09/894,912A  
PRIORITY FILING DATE: 2002-05-10  
PRIORITY APPLICATION NUMBER: To be assigned  
PRIORITY FILING DATE: 2001-04-05  
PRIORITY APPLICATION NUMBER: 60/266,614  
PRIORITY FILING DATE: 2001-02-05  
PRIORITY APPLICATION NUMBER: 60/215,733  
PRIORITY FILING DATE: 2000-06-28  
PRIORITY APPLICATION NUMBER: 09/757,562  
PRIORITY FILING DATE: 2001-01-09  
PRIORITY APPLICATION NUMBER: 09/543,774  
PRIORITY FILING DATE: 2000-04-05  
NUMBER OF SEQ ID NOS: 48  
SOFTWARE: PatentIn version 3.0  
SEQ ID NO 25  
LENGTH: 229  
TYPE: PRT  
ORGANISM: Mus musculus  
US-09-894-912A-25

Query Match 45.0%; Score 40.5; DB 9; Length 229;  
Best Local Similarity 56.2%; Pred. No. 32;  
Matches 9; Conservative 3; Mismatches 3; Indels 1; Gaps 1;

QY 2 GYGRDNDNRPSRKN 16  
: : : : :  
Db 198 GQGRRENANRHPARKN 213

RESULT 8  
US-10-125-852-24  
Sequence 24, Application US/10125852  
Publication No. US20030032034A1  
GENERAL INFORMATION:  
APPLICANT: Tang, Y. Tom  
TITLE OF INVENTION: METHODS AND MATERIALS RELATING TO NOVEL STEM CELL GROWTH FACTOR-1  
FILE REFERENCE: HYS-43A  
CURRENT APPLICATION NUMBER: US/10/125,852  
PRIORITY FILING DATE: 2002-08-20  
PRIORITY APPLICATION NUMBER: US 60/316,368  
PRIORITY FILING DATE: 2001-08-30  
PRIORITY APPLICATION NUMBER: US 09/799,451  
PRIORITY FILING DATE: 2001-03-05  
NUMBER OF SEQ ID NOS: 25  
SOFTWARE: PatentIn version 3.1  
SEQ ID NO 24  
LENGTH: 265  
TYPE: PRT  
ORGANISM: Mus musculus  
US-10-125-852-24

Query Match 45.0%; Score 40.5; DB 9; Length 265;  
Best Local Similarity 56.2%; Pred. No. 38;  
Matches 9; Conservative 3; Mismatches 3; Indels 1; Gaps 1;

QY 2 GYGRDNDNRPSRKN 16  
: : : : :  
Db 216 GQGRRENANRHPARKN 231

RESULT 9  
US-09-815-242-11216  
Sequence 11216, Application US/09815242  
Patent No. US20020061569A1  
GENERAL INFORMATION:  
APPLICANT: Haselbeck, Robert  
APPLICANT: Ohlser, Karl L.  
APPLICANT: Zyckind, Judith W.  
APPLICANT: Wall, Daniel  
APPLICANT: Trawick, John D.  
APPLICANT: Carr, Grant J.  
APPLICANT: Yamamoto, Robert T.  
APPLICANT: Xu, H. Howard  
TITLE OF INVENTION: Identification of Essential Genes in  
FILE REFERENCE: ELITRA.011A  
CURRENT APPLICATION NUMBER: US/09/815,242  
PRIORITY FILING DATE: 2001-03-21  
PRIORITY APPLICATION NUMBER: 60/191,078  
PRIORITY FILING DATE: 2000-03-21  
PRIORITY APPLICATION NUMBER: 60/206,848  
PRIORITY FILING DATE: 2000-05-23  
PRIORITY APPLICATION NUMBER: 60/207,727  
PRIORITY FILING DATE: 2000-05-26  
PRIORITY APPLICATION NUMBER: 60/242,578  
PRIORITY FILING DATE: 2000-10-23  
PRIORITY APPLICATION NUMBER: 60/253,625  
PRIORITY FILING DATE: 2000-11-27  
PRIORITY APPLICATION NUMBER: 60/257,931  
PRIORITY FILING DATE: 2000-12-22  
PRIORITY APPLICATION NUMBER: 60/269,308  
PRIORITY FILING DATE: 2001-02-16  
NUMBER OF SEQ ID NOS: 14110  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 11216  
LENGTH: 829  
TYPE: PRT  
ORGANISM: Haemophilus influenzae  
US-09-815-242-11216

Query Match 45.0%; Score 40.5; DB 10; Length 829;  
Best Local Similarity 45.0%; Pred. No. 1.3e+02;  
Matches 9; Conservative 3; Mismatches 1; Indels 7; Gaps 1;

QY 4 GRDD-----NDRPSRKN 16  
: : : : :  
Db 197 GRDDNSKNSKNERESNRKN 216

RESULT 10  
US-09-866-050A-684  
Sequence 684, Application US/09866050A  
Publication No. US20030040471A1  
GENERAL INFORMATION:  
APPLICANT: Watson, James D.  
APPLICANT: Strachan, Lorna  
APPLICANT: Sleeman, Matthew  
APPLICANT: Onrust, Rene  
APPLICANT: Marison, James G.  
APPLICANT: Kumble, Krishanand D.  
TITLE OF INVENTION: Compositions Isolated From Skin Cells  
FILE REFERENCE: 11000.1011c4U  
CURRENT APPLICATION NUMBER: US/09/866,050A

```
; CURRENT FILING DATE: 2001-05-24
; NUMBER OF SEQ ID NOS: 725
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 684
; LENGTH: 228
; TYPE: PRF
; ORGANISM: Mouse
US-09-866-050A-684

Query Match      44.4%; Score 40; DB 9; Length 228;
Best Local Similarity 40.0%; Pred. No. 39;
Matches 6; Conservative 4; Mismatches 5; Indels 0; Gaps 0;

QY      2 GYGRDNDNDPSRKX 16
       |::|||::|
Db      20 GMAQEDGDADPEPEN 34

RESULT 11
US-09-870-759-90
; Sequence 90, Application US/09870759
; Patent No. US20020177551A1
; GENERAL INFORMATION:
; APPLICANT: TERMAN, David S
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR TREATMENT OF NEOPLASTIC DISEASE
; FILE REFERENCE: 870759
; CURRENT APPLICATION NUMBER: US/09/870,759
; CURRENT FILING DATE: 2002-01-14
; PRIOR APPLICATION NUMBER: US 60/208,128
; PRIOR FILING DATE: 2000-05-30
; NUMBER OF SEQ ID NOS: 166
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 90
; LENGTH: 104
; TYPE: PRF
; ORGANISM: Staphylococcus aureus
US-09-870-759-90

Query Match      43.3%; Score 39; DB 9; Length 104;
Best Local Similarity 60.0%; Pred. No. 24;
Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY      6 DDNDPSPSRK 15
       ||::|||
Db      27 DDDKDPSSR 36

RESULT 12
US-10-083-590-12
; Sequence 12, Application US/10083590
; Publication No. US20030027257A1
; GENERAL INFORMATION:
; APPLICANT: IATROU, Kostas
; APPLICANT: FARRELL, Patrick J.
; TITLE OF INVENTION: SEQUENCES FOR IMPROVING THE EFFICIENCY OF SECRETION OF
; TITLE OF INVENTION: NON-SECRETED PROTEINS FROM MAMMALIAN AND INSECT CELLS
; FILE REFERENCE: 028722-207
; CURRENT APPLICATION NUMBER: US/10/083,590
; CURRENT FILING DATE: 2002-02-27
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 09/256,694
; PRIOR FILING DATE: EARLIER FILING DATE: 1999-02-24
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: US 09/136,421
; PRIOR FILING DATE: EARLIER FILING DATE: 1998-08-20
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: US 60/056,871
; PRIOR FILING DATE: EARLIER FILING DATE: 1997-08-21
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 12
; LENGTH: 17
; TYPE: PRF
; ORGANISM: Artificial Sequence
; FEATURE:
```

```
; OTHER INFORMATION: Description of Artificial Sequence: Has a cleavage
; OTHER INFORMATION: site recognized by the protease porcine intestine
; OTHER INFORMATION: enteropeptidase.
US-10-083-590-12

Query Match      42.2%; Score 38; DB 9; Length 17;
Best Local Similarity 66.7%; Pred. No. 5;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY      4 GRDNDNDP 12
       ||::||
Db      9 GDDDDKDP 17

RESULT 13
US-09-925-287-1
; Sequence 1, Application US/09925287
; Publication No. US20030031653A1
; GENERAL INFORMATION:
; APPLICANT: FARRELL, Patrick J.
; APPLICANT: GEDAMU, Lashitew
; APPLICANT: IATROU, Kostas
; TITLE OF INVENTION: Compositions and Methods for Eliciting Immune Responses
; FILE REFERENCE: 028722-275
; CURRENT APPLICATION NUMBER: US/09/925,287
; CURRENT FILING DATE: 2001-08-08
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1
; LENGTH: 17
; TYPE: PRF
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: amino acid sequence of the spacer used
US-09-925-287-1

Query Match      42.2%; Score 38; DB 9; Length 17;
Best Local Similarity 66.7%; Pred. No. 5;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY      4 GRDNDNDP 12
       ||::||
Db      9 GDDDDKDP 17

RESULT 14
US-10-083-815-68
; Sequence 68, Application US/10083815
; Publication No. US20030026781A1
; GENERAL INFORMATION:
; APPLICANT: Anderson, Christen M.
; APPLICANT: CLEVENGER, William
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR REGULATING
; TITLE OF INVENTION: ENDOGENOUS INHIBITOR OF ATP SYNTHASE, INCLUDING
; FILE REFERENCE: 660088.435C2
; CURRENT APPLICATION NUMBER: US/10/083,815
; CURRENT FILING DATE: 2002-02-27
; NUMBER OF SEQ ID NOS: 72
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 68
; LENGTH: 35
; TYPE: PRF
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Epitope tag sequence.
US-10-083-815-68

Query Match      42.2%; Score 38; DB 9; Length 35;
Best Local Similarity 75.0%; Pred. No. 11;
Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;
```



OY 6 DDNDPDS 13  
Db 27 DDDKDPDS 34

RESULT 15  
US-10-083-815-71

; Sequence 71, Application US/10083815  
; Publication No. US20030026781A1  
; GENERAL INFORMATION:  
; APPLICANT: Anderson, Christen M.  
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR REGULATING  
; TITLE OF INVENTION: ENDOGENOUS INHIBITOR OF ATP SYNTHASE, INCLUDING  
; FILE REFERENCE: 660088.435C2  
; CURRENT APPLICATION NUMBER: US/10/083,815  
; NUMBER OF SEQ ID NOS: 72  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 71  
; LENGTH: 107  
; TYPE: PRT  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Fusion protein  
US-10-083-815-71

Query Match 42.2%; Score 38; DB 9; Length 107;  
Best Local Similarity 75.0%; Pred. No. 36;  
Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

OY 6 DDNDPDS 13  
Db 27 DDDKDPDS 34

RESULT 16  
US-09-864-761-35255

; Sequence 35255, Application US/09864761  
; Patent No. US20020048763A1  
; GENERAL INFORMATION:  
; APPLICANT: Penn, Sharon G.  
; APPLICANT: Rank, David R.  
; APPLICANT: Hanzel, David K.  
; APPLICANT: Chen, Wensheng  
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR  
; FILE REFERENCE: Aesomica-X-1  
; CURRENT APPLICATION NUMBER: US/09/864,761  
; CURRENT FILING DATE: 2001-05-23  
; PRIOR APPLICATION NUMBER: US 60/180,312  
; PRIOR FILING DATE: 2000-02-04  
; PRIOR APPLICATION NUMBER: US 60/207,456  
; PRIOR FILING DATE: 2000-05-26  
; PRIOR APPLICATION NUMBER: US 09/632,366  
; PRIOR FILING DATE: 2000-08-03  
; PRIOR APPLICATION NUMBER: GB 24263,6  
; PRIOR FILING DATE: 2000-10-04  
; PRIOR APPLICATION NUMBER: US 60/236,359  
; PRIOR FILING DATE: 2000-09-27  
; PRIOR APPLICATION NUMBER: PCT/US01/00666  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00667  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00664  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00669  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00665  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00668  
; PRIOR FILING DATE: 2001-01-30

; PRIOR APPLICATION NUMBER: PCT/US01/00663  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00662  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00661  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00670  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: US 60/234,687  
; PRIOR FILING DATE: 2000-09-21  
; PRIOR APPLICATION NUMBER: US 09/608,408  
; PRIOR FILING DATE: 2000-06-30  
; PRIOR APPLICATION NUMBER: US 09/774,203  
; PRIOR FILING DATE: 2001-01-29  
; NUMBER OF SEQ ID NOS: 49117  
; SOFTWARE: Annonmax Sequence Listing Engine vers. 1.1  
; SEQ ID NO 35255  
; LENGTH: 176  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
; FEATURE:  
; OTHER INFORMATION: MAP TO AL035660.11  
; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 1  
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 1.1  
; OTHER INFORMATION: EXPRESSED IN HBL100, SIGNAL = 1.4  
; OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 0.92  
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 1.6  
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 1.4  
; OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 1.5  
; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 0.92  
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 1.1  
; OTHER INFORMATION: SWISSPROT HIT: P04279, EVALUATE 1.00e-102  
; OTHER INFORMATION: EST\_HUMAN HIT: BF679236.1, EVALUATE 1.00e-100

US-09-864-761-35255

Query Match 42.2%; Score 38; DB 10; Length 176;  
Best Local Similarity 57.1%; Pred. No. 61;  
Matches 8; Conservative 1; Mismatches 5; Indels 0; Gaps 0;

OY 3 YGRDNDPDSRXN 16  
Db 43 YHVDNDHDSRKS 56

RESULT 17  
US-09-925-301-983

; Sequence 983, Application US/09925301  
; Patent No. US20020052308A1  
; GENERAL INFORMATION:  
; APPLICANT: Rosen et al.  
; TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies  
; FILE REFERENCE: PA106  
; CURRENT APPLICATION NUMBER: US/09/925,301  
; CURRENT FILING DATE: 2001-08-10  
; PRIOR APPLICATION NUMBER: PCT/US00/05882  
; PRIOR FILING DATE: 2000-03-08  
; PRIOR APPLICATION NUMBER: 60/124,270  
; PRIOR FILING DATE: 1999-03-12  
; NUMBER OF SEQ ID NOS: 1694  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 983  
; LENGTH: 193  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: SITE  
; LOCATION: (72)  
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids  
; NAME/KEY: SITE  
; LOCATION: (135)  
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids  
; NAME/KEY: SITE  
; LOCATION: (139)

; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids  
US-09-925-301-983

Query Match 42.2%; Score 38; DB 10; Length 193;  
Best Local Similarity 61.5%; Pred. No. 68;  
Matches 8; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 2 GYGRDNDNDPSR 14  
|||  
Db 148 GYPRDRNSGVPSR 160

RESULT 18

US-09-286-488-38  
; Sequence 38, Application US/09286488  
; Patent No. US20020169136A1  
; GENERAL INFORMATION:  
; APPLICANT: Reed, Steven G.  
; APPLICANT: Lodes, Michael J.  
; APPLICANT: Houghton, Raymond L.  
; APPLICANT: Sleath, Paul R.  
; APPLICANT: McNeill, Patricia D.  
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THE DIAGNOSIS  
; FILE REFERENCE: 210121.426C3  
; CURRENT APPLICATION NUMBER: US/09/286.488  
; CURRENT FILING DATE: 1999-04-05  
; NUMBER OF SEQ ID NOS: 83  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 38  
; LENGTH: 445  
; TYPE: PRT  
; ORGANISM: Babesia microti  
US-09-286-488-38

Query Match 42.2%; Score 38; DB 9; Length 445;  
Best Local Similarity 75.0%; Pred. No. 1.7e+02;  
Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 6 DDNDRDPS 13  
:|:|:|  
Db 381 EDNERDPS 388

RESULT 19

US-09-737-178-38  
; Sequence 38, Application US/09737178  
; Patent No. US20010029295A1  
; GENERAL INFORMATION:  
; APPLICANT: Reed, Steven G.  
; APPLICANT: Lodes, Michael J.  
; APPLICANT: Houghton, Raymond L.  
; APPLICANT: Sleath, Paul R.  
; APPLICANT: McNeill, Patricia D.  
; APPLICANT: Homer, Mary  
; APPLICANT: Secrist, Heather  
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THE DIAGNOSIS  
; FILE REFERENCE: 210121.426C9  
; CURRENT APPLICATION NUMBER: US/09/737.178  
; CURRENT FILING DATE: 2000-12-13  
; NUMBER OF SEQ ID NOS: 144  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 38  
; LENGTH: 445  
; TYPE: PRT  
; ORGANISM: Babesia microti  
US-09-737-178-38

Query Match 42.2%; Score 38; DB 10; Length 445;  
Best Local Similarity 75.0%; Pred. No. 1.7e+02;  
Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 6 DDNDRDPS 13  
:|:|:|  
Db 381 EDNERDPS 388

RESULT 20

US-09-738-626-4373  
; Sequence 4373, Application US/09738626  
; Publication No. US20020197605A1  
; GENERAL INFORMATION:  
; APPLICANT: NAKAGAWA, SATOSHI  
; APPLICANT: MIZOGUCHI, HIROSHI  
; APPLICANT: ANDO, SEIKO  
; APPLICANT: HAYASHI, MIKIRO  
; APPLICANT: OCHIAI, KEIKO  
; APPLICANT: YOKOI, HARUHIKO  
; APPLICANT: TATEISHI, NAOKO  
; APPLICANT: SENOH, AKIHIRO  
; APPLICANT: IKEDA, MASATO  
; APPLICANT: OZAKI, AKIO  
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES  
; FILE REFERENCE: 249-125  
; CURRENT APPLICATION NUMBER: US/09/738.626  
; CURRENT FILING DATE: 2000-12-18  
; PRIOR APPLICATION NUMBER: JP 99/377484  
; PRIOR FILING DATE: 1999-12-16  
; PRIOR APPLICATION NUMBER: JP 00/159162  
; PRIOR FILING DATE: 2000-04-07  
; PRIOR APPLICATION NUMBER: JP 00/280988  
; PRIOR FILING DATE: 2000-08-03  
; NUMBER OF SEQ ID NOS: 7059  
; SOFTWARE: PatentIn ver. 3.0  
; SEQ ID NO 4373  
; LENGTH: 482  
; TYPE: PRT  
; ORGANISM: Corynebacterium glutamicum  
US-09-738-626-4373

Query Match 42.2%; Score 38; DB 9; Length 482;  
Best Local Similarity 42.9%; Pred. No. 1.8e+02;  
Matches 6; Conservative 4; Mismatches 4; Indels 0; Gaps 0;

QY 3 YGRDNDNDPSRKN 16  
:|:|:|:|  
Db 17 FGNNDDDDKXNN 30

RESULT 21

US-10-138-838-95  
; Sequence 95, Application US/10138838  
; Publication No. US20030049821A1  
; GENERAL INFORMATION:  
; APPLICANT: Wilson, Ron C.  
; APPLICANT: Craft, David L.  
; APPLICANT: Birch, Dudley  
; APPLICANT: Eshoo, Mark  
; APPLICANT: Madduri, Krishna M.  
; APPLICANT: Cornett, Cathy A.  
; APPLICANT: Brenner, Alfred A.  
; APPLICANT: Tang, Maria  
; APPLICANT: Loper, John C.  
; APPLICANT: Gleeson, Martin  
; TITLE OF INVENTION: CYTOCHROME P450 OXIDOREDUCTASE  
; TITLE OF INVENTION: P450 OXIDOREDUCTASE  
; TITLE OF INVENTION: GENES AND PROTEINS RELATED TO THE OMEGA HYDROXYLASE COMPLEX OF  
; FILE REFERENCE: 1010-16  
; CURRENT APPLICATION NUMBER: US/10/138.838  
; CURRENT FILING DATE: 2002-05-03  
; PRIOR APPLICATION NUMBER: US/09/976.800  
; PRIOR FILING DATE: 2001-10-12  
; NUMBER OF SEQ ID NOS: 118  
; SOFTWARE: PatentIn version 3.1

SEQ ID NO 95  
LENGTH: 523  
TYPE: PRT  
ORGANISM: CANDIDATROPICALIS  
US-10-138-838-95

Query Match 42.2%; Score 38; DB 9; Length 523;  
Best Local Similarity 58.3%; Pred. No. 2e+02;  
Matches 7; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

OY 3 YGRDNDRDPGR 14  
DB 437 YGKDANDRPRR 448

RESULT 22  
US-10-138-838-110  
Sequence 110, Application US/10138938  
Publication No. US20030049822A1

GENERAL INFORMATION:  
APPLICANT: Wilson, Ron C.  
APPLICANT: Craft, David L.  
APPLICANT: Eitich, Dudley  
APPLICANT: Eshoo, Mark  
APPLICANT: Madduri, Krishna M.  
APPLICANT: Cornett, Cathy A.  
APPLICANT: Brenner, Alfred A.  
APPLICANT: Tang, Maria  
APPLICANT: Loper, John C.  
APPLICANT: Gleeson, Martin  
TITLE OF INVENTION: CYTOCHROME P450 MONOOXYGENASE AND NADPH CYTOCHROME  
TITLE OF INVENTION: P450 OXIDOREDUCTASE  
TITLE OF INVENTION: GENES AND PROTEINS RELATED TO THE OMEGA HYDROXYLASE COMPLEX OF  
TITLE OF INVENTION: TROPICALIS AND METHODS RELATING THERETO  
FILE REFERENCE: 1010-16  
CURRENT APPLICATION NUMBER: US/10/138, 838  
CURRENT FILING DATE: 2002-05-03  
PRIOR APPLICATION NUMBER: US/09/976, 800  
PRIOR FILING DATE: 2001-10-12  
NUMBER OF SEQ ID NOS: 118  
SOFTWARE: PatentIn version 3.1  
SEQ ID NO 110  
LENGTH: 523  
TYPE: PRT  
ORGANISM: CANDIDATROPICALIS  
US-10-138-838-110

Query Match 42.2%; Score 38; DB 9; Length 523;  
Best Local Similarity 58.3%; Pred. No. 2e+02;  
Matches 7; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

OY 3 YGRDNDRDPGR 14  
DB 437 YGKDANDRPRR 448

RESULT 23  
US-10-139-031-95  
Sequence 95, Application US/10139031  
Publication No. US20030049822A1

GENERAL INFORMATION:  
APPLICANT: Wilson, Ron C.  
APPLICANT: Craft, David L.  
APPLICANT: Eitich, Dudley  
APPLICANT: Eshoo, Mark  
APPLICANT: Madduri, Krishna M.  
APPLICANT: Cornett, Cathy A.  
APPLICANT: Brenner, Alfred A.  
APPLICANT: Tang, Maria  
APPLICANT: Loper, John C.  
APPLICANT: Gleeson, Martin  
TITLE OF INVENTION: CYTOCHROME P450 MONOOXYGENASE AND NADPH CYTOCHROME  
TITLE OF INVENTION: P450 OXIDOREDUCTASE

TITLE OF INVENTION: GENES AND PROTEINS RELATED TO THE OMEGA HYDROXYLASE COMPLEX OF  
TITLE OF INVENTION: TROPICALIS AND METHODS RELATING THERETO  
FILE REFERENCE: 1010-16  
CURRENT APPLICATION NUMBER: US/10/139, 031  
CURRENT FILING DATE: 2002-05-03  
PRIOR APPLICATION NUMBER: US/09/976, 800  
PRIOR FILING DATE: 2001-10-12  
NUMBER OF SEQ ID NOS: 118  
SOFTWARE: PatentIn version 3.1  
SEQ ID NO 95  
LENGTH: 523  
TYPE: PRT  
ORGANISM: CANDIDATROPICALIS  
US-10-139-031-95

Query Match 42.2%; Score 38; DB 9; Length 523;  
Best Local Similarity 58.3%; Pred. No. 2e+02;  
Matches 7; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

OY 3 YGRDNDRDPGR 14  
DB 437 YGKDANDRPRR 448

RESULT 24  
US-10-139-031-110  
Sequence 110, Application US/10139031  
Publication No. US20030049822A1

GENERAL INFORMATION:  
APPLICANT: Wilson, Ron C.  
APPLICANT: Craft, David L.  
APPLICANT: Eitich, Dudley  
APPLICANT: Eshoo, Mark  
APPLICANT: Madduri, Krishna M.  
APPLICANT: Cornett, Cathy A.  
APPLICANT: Brenner, Alfred A.  
APPLICANT: Tang, Maria  
APPLICANT: Loper, John C.  
APPLICANT: Gleeson, Martin  
TITLE OF INVENTION: CYTOCHROME P450 MONOOXYGENASE AND NADPH CYTOCHROME  
TITLE OF INVENTION: P450 OXIDOREDUCTASE  
TITLE OF INVENTION: GENES AND PROTEINS RELATED TO THE OMEGA HYDROXYLASE COMPLEX OF  
TITLE OF INVENTION: TROPICALIS AND METHODS RELATING THERETO  
FILE REFERENCE: 1010-16  
CURRENT APPLICATION NUMBER: US/10/139, 031  
CURRENT FILING DATE: 2002-05-03  
PRIOR APPLICATION NUMBER: US/09/976, 800  
PRIOR FILING DATE: 2001-10-12  
NUMBER OF SEQ ID NOS: 118  
SOFTWARE: PatentIn version 3.1  
SEQ ID NO 110  
LENGTH: 523  
TYPE: PRT  
ORGANISM: CANDIDATROPICALIS  
US-10-139-031-110

Query Match 42.2%; Score 38; DB 9; Length 523;  
Best Local Similarity 58.3%; Pred. No. 2e+02;  
Matches 7; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

OY 3 YGRDNDRDPGR 14  
DB 437 YGKDANDRPRR 448

RESULT 25  
US-09-737-178-85  
Sequence 85, Application US/09737178  
Patent No. US20010029295A1

GENERAL INFORMATION:  
APPLICANT: Reed, Steven G.  
APPLICANT: Lodes, Michael J.  
APPLICANT: Houghton, Raymond L.

```
; APPLICANT: Sleath, Paul R.
; APPLICANT: McNeill, Patricia D.
; APPLICANT: Homer, Mary
; APPLICANT: Secret, Heather
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THE DIAGNOSIS
; FILE REFERENCE: 210121.426C9
; CURRENT APPLICATION NUMBER: US/09/737,178
; CURRENT FILING DATE: 2000-12-13
; NUMBER OF SEQ ID NOS: 144
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 85
; LENGTH: 666
; TYPE: PRT
; ORGANISM: Babesia
US-09-737-178-85

Query Match          42.2%; Score 38; DB 10; Length 666;
Best Local Similarity 75.0%; Pred. No. 2.6e+02;
Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY      6 DDNDRDPS 13
Db      602 EDNERDPS 609

RESULT 26
US-09-737-178-144
; Sequence 144, Application US/09737178
; Patent No. US20010029295A1
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Lodes, Michael J.
; APPLICANT: Houghton, Raymond L.
; APPLICANT: Sleath, Paul R.
; APPLICANT: McNeill, Patricia D.
; APPLICANT: Homer, Mary
; APPLICANT: Secret, Heather
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THE DIAGNOSIS
; FILE REFERENCE: 210121.426C9
; CURRENT APPLICATION NUMBER: US/09/737,178
; CURRENT FILING DATE: 2000-12-13
; NUMBER OF SEQ ID NOS: 144
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 144
; LENGTH: 677
; TYPE: PRT
; ORGANISM: Babesia
US-09-737-178-144

Query Match          42.2%; Score 38; DB 10; Length 677;
Best Local Similarity 75.0%; Pred. No. 2.6e+02;
Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY      6 DDNDRDPS 13
Db      374 EDNERDPS 381

RESULT 27
US-09-737-178-87
; Sequence 87, Application US/09737178
; Patent No. US20010029295A1
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Lodes, Michael J.
; APPLICANT: Houghton, Raymond L.
; APPLICANT: Sleath, Paul R.
; APPLICANT: McNeill, Patricia D.
; APPLICANT: Homer, Mary
; APPLICANT: Secret, Heather
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THE DIAGNOSIS
```

```
; TITLE OF INVENTION: AND TREATMENT OF B. MICROTI INFECTION
; FILE REFERENCE: 210121.426C9
; CURRENT APPLICATION NUMBER: US/09/737,178
; CURRENT FILING DATE: 2000-12-13
; NUMBER OF SEQ ID NOS: 144
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 87
; LENGTH: 1132
; TYPE: PRT
; ORGANISM: Babesia
US-09-737-178-87

Query Match          42.2%; Score 38; DB 10; Length 1132;
Best Local Similarity 75.0%; Pred. No. 4.3e+02;
Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY      6 DDNDRDPS 13
Db      1068 EDNERDPS 1075

RESULT 28
US-09-864-761-36056
; Sequence 36056, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharon G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
; FILE REFERENCE: Aeomica-X-1
; CURRENT APPLICATION NUMBER: US/09/864,761
; CURRENT FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/632,366
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: GB 24263.6
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US01/00666
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00664
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00665
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00668
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00663
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00662
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00661
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00670
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: US 09/608,408
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/774,203
; PRIOR FILING DATE: 2001-01-29
; NUMBER OF SEQ ID NOS: 49117
; SOFTWARE: Anomax Sequence Listing Engine vers. 1.1
```

```
; SEQ ID NO 36056
; LENGTH: 31
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AC008125.9
; OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 2.6
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 6.5
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 2.9
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 6.5
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 5.8
; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 1.8
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 5
; OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 1.1
; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 3.8
; OTHER INFORMATION: EXPRESSED IN HBL100, SIGNAL = 3
US-09-864-761-36056
```

```
Query Match 41.1%; Score 37; DB 10; Length 31;
Best Local Similarity 70.0%; Pred. No. 14;
Matches 7; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
```

```
OY 2 GYGDDNDRD 11
Db 4 GYGDDADGD 13
```

## RESULT 29

```
US-09-864-761-35018
; Sequence 35018, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharon G.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
; FILE REFERENCE: Aecm1ca-X-1
; CURRENT APPLICATION NUMBER: US/09/864,761
; CURRENT FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/632,366
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: GB 24263,6
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US01/00666
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00664
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00669
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00665
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00668
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00663
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00662
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00661
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00670
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 2000-09-21
```

```
; PRIOR APPLICATION NUMBER: US 09/608,408
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/774,203
; PRIOR FILING DATE: 2001-01-29
; NUMBER OF SEQ ID NOS: 49117
; SOFTWARE: Annonmax Sequence Listing Engine vers. 1.1
; SEQ ID NO 35018
; LENGTH: 42
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
```

```
; OTHER INFORMATION: MAP TO AL049553.14
; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 2.3
; OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 2.4
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 2.6
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 2.8
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 1.2
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 1.9
; OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 2.4
; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 3
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 2.5
; OTHER INFORMATION: EXPRESSED IN HBL100, SIGNAL = 2.4
US-09-864-761-35018
```

```
Query Match 41.1%; Score 37; DB 10; Length 42;
Best Local Similarity 70.0%; Pred. No. 19;
Matches 7; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
```

```
OY 6 DDNDRDPGRK 15
Db 21 DDNDDDGRK 30
```

## RESULT 30

```
US-09-983-802-172
; Sequence 172, Application US/09983802
; Publication No. US20030022185A1
; GENERAL INFORMATION:
; APPLICANT: Fischer et al.
; TITLE OF INVENTION: 123 Human Secreted Proteins
; FILE REFERENCE: P2010P1
; CURRENT APPLICATION NUMBER: US/09/983,802
; CURRENT FILING DATE: 2001-10-25
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 09/227,357
; PRIOR FILING DATE: EARLIER FILING DATE: 1999-01-08
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: PCT/US98/13684
; PRIOR FILING DATE: EARLIER FILING DATE: 1998-07-07
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/051,926
; PRIOR FILING DATE: EARLIER FILING DATE: 1997-07-08
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/052,793
; PRIOR FILING DATE: EARLIER FILING DATE: 1997-07-08
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/051,925
; PRIOR FILING DATE: EARLIER FILING DATE: 1997-07-08
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/051,929
; PRIOR FILING DATE: EARLIER FILING DATE: 1997-07-08
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/052,803
; PRIOR FILING DATE: EARLIER FILING DATE: 1997-07-08
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/052,732
; PRIOR FILING DATE: EARLIER FILING DATE: 1997-07-08
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/051,931
; PRIOR FILING DATE: EARLIER FILING DATE: 1997-07-08
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/051,932
; PRIOR FILING DATE: EARLIER FILING DATE: 1997-07-08
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/051,916
; PRIOR FILING DATE: EARLIER FILING DATE: 1997-07-08
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/051,930
; PRIOR FILING DATE: EARLIER FILING DATE: 1997-07-08
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/051,918
; PRIOR FILING DATE: EARLIER FILING DATE: 1997-07-08
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/051,920
; PRIOR FILING DATE: EARLIER FILING DATE: 1997-07-08
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/052,733
; PRIOR FILING DATE: EARLIER FILING DATE: 1997-07-08
```

```
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/052,795
; PRIOR FILING DATE: EARLIER FILING DATE: 1997-07-08
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/051,919
; PRIOR FILING DATE: EARLIER FILING DATE: 1997-07-08
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/051,928
; PRIOR FILING DATE: EARLIER FILING DATE: 1997-07-08
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/055,722
; PRIOR FILING DATE: EARLIER FILING DATE: 1997-08-18
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/055,723
; PRIOR FILING DATE: EARLIER FILING DATE: 1997-08-18
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/055,948
; PRIOR FILING DATE: EARLIER FILING DATE: 1997-08-18
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/055,949
; PRIOR FILING DATE: EARLIER FILING DATE: 1997-08-18
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/055,953
; PRIOR FILING DATE: EARLIER FILING DATE: 1997-08-18
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/055,950
; PRIOR FILING DATE: EARLIER FILING DATE: 1997-08-18
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/055,947
; PRIOR FILING DATE: EARLIER FILING DATE: 1997-08-18
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/055,964
; PRIOR FILING DATE: EARLIER FILING DATE: 1997-08-18
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/055,684
; PRIOR FILING DATE: EARLIER FILING DATE: 1997-08-18
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/056,360
; PRIOR FILING DATE: EARLIER FILING DATE: 1997-08-18
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/055,984
; PRIOR FILING DATE: EARLIER FILING DATE: 1997-08-18
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/058,785
; PRIOR FILING DATE: EARLIER FILING DATE: 1997-09-12
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/058,664
; PRIOR FILING DATE: EARLIER FILING DATE: 1997-09-12
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/058,660
; PRIOR FILING DATE: EARLIER FILING DATE: 1997-09-12
; PRIOR APPLICATION NUMBER: EARLIER FILING DATE: 1997-09-12
; PRIOR FILING DATE: EARLIER FILING DATE: 1997-09-12
; NUMBER OF SEQ ID NOS: 672
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 172
; LENGTH: 104
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-983-802-172

Query Match          41.1%; Score 37; DB 9; Length 104;
Best Local Similarity 50.0%; Pred. NO. 51;
Matches 6; Conservative 3; Mismatches 3; Indels 0; Gaps 0;
```

```
Qy      3 YGRDNDRDPSPR 14
       :|:|:|:|
Db      93 WGDQDQEHVPSR 104
```

```
RESULT 31
US-09-815-242-10900
; Sequence 10900, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari L.
; APPLICANT: Zysek, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; TITLE OF INVENTION: Prokaryotes
; FILE REFERENCE: EUTRA.011a
; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21
```

```
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 1410
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 10900
; LENGTH: 333
; TYPE: PRT
; ORGANISM: Enterococcus faecalis
US-09-815-242-10900

Query Match          41.1%; Score 37; DB 10; Length 333;
Best Local Similarity 46.7%; Pred. NO. 1.8e+02;
Matches 7; Conservative 2; Mismatches 6; Indels 0; Gaps 0;
```

```
Qy      2 GYGRDNDRDPSPRN 16
       :|:|:|:|
Db      264 GIGTDVLEBRGRKN 278
```

```
RESULT 32
US-09-815-242-12046
; Sequence 12046, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari L.
; APPLICANT: Zysek, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; TITLE OF INVENTION: Prokaryotes
; FILE REFERENCE: EUTRA.011a
; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 1410
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 12046
; LENGTH: 774
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-815-242-12046

Query Match          41.1%; Score 37; DB 10; Length 774;
Best Local Similarity 58.3%; Pred. NO. 4.4e+02;
```

Matches 7; Conservative 1; Mismatches 4; Indels 0; Gaps 0;  
Qy 1 VGYGRDDNDP 12  
Db 677 VMLGRDDNGKTP 688

RESULT 33  
US-09-801-368-324  
; Sequence 324, Application US/09801368  
; Patent No. US20020128250A1  
; GENERAL INFORMATION:  
; APPLICANT: Busby, Robert  
; APPLICANT: Cali, Brian  
; APPLICANT: Hecht, Peter  
; APPLICANT: Holtzman, Doug  
; APPLICANT: Madden, Kevin  
; APPLICANT: Maxon, Mary  
; APPLICANT: Milne, Todd  
; APPLICANT: No. US20020128250A1man, Thea  
; APPLICANT: Royer, John  
; APPLICANT: Salama, Sofie  
; APPLICANT: Sherman, Amit  
; APPLICANT: Silva, Jeff  
; APPLICANT: Summers, Eric  
; TITLE OF INVENTION: Methods for Improving Secondary Metabolite Production in Fungi  
; FILE REFERENCE: 109272.147  
; CURRENT APPLICATION NUMBER: US/09/801,368  
; CURRENT FILING DATE: 2001-03-07  
; PRIOR APPLICATION NUMBER: US 09/487,558  
; PRIOR FILING DATE: 2000-01-19  
; PRIOR APPLICATION NUMBER: US 60/160,587  
; PRIOR FILING DATE: 1999-10-20  
; NUMBER OF SEQ ID NOS: 440  
; SOFTWARE: PatentIn version 3.0  
; SEQ ID NO 324  
; LENGTH: 829  
; TYPE: PRT  
; ORGANISM: *Saccharomyces cerevisiae*  
US-09-801-368-324

Query Match 41.1%; Score 37; DB 10; Length 829;  
Best Local Similarity 40.0%; Pred. No. 4.7e+02;  
Matches 6; Conservative 2; Mismatches 7; Indels 0; Gaps 0;

Qy 2 GYGRDDNDPSPK 16  
Db 722 GYAEDDDEDSN 736

RESULT 34  
US-09-738-626-5676  
; Sequence 5676, Application US/09738626  
; Publication No. US20020197605A1  
; GENERAL INFORMATION:  
; APPLICANT: NAKAGAWA, SATOSHI  
; APPLICANT: MIZOGUCHI, HIROSHI  
; APPLICANT: ANDO, SEIKO  
; APPLICANT: HAYASHI, MIKIRO  
; APPLICANT: OCHIAI, KEIKO  
; APPLICANT: YOKOI, HARUHIKO  
; APPLICANT: TATEISHI, MOKO  
; APPLICANT: SENOH, AKIHIRO  
; APPLICANT: IKEDA, MASATO  
; APPLICANT: OZAKI, AKIO  
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES  
; FILE REFERENCE: 249-125  
; CURRENT APPLICATION NUMBER: US/09/738,626  
; CURRENT FILING DATE: 2000-12-18  
; PRIOR APPLICATION NUMBER: JP 99/377484  
; PRIOR FILING DATE: 1999-12-16  
; PRIOR APPLICATION NUMBER: JP 00/159162  
; PRIOR FILING DATE: 2000-04-07

; PRIOR APPLICATION NUMBER: JP 00/280988  
; PRIOR FILING DATE: 2000-08-03  
; NUMBER OF SEQ ID NOS: 7059  
; SOFTWARE: PatentIn ver. 3.0  
; SEQ ID NO 5676  
; LENGTH: 1004  
; TYPE: PRT  
; ORGANISM: *Corynebacterium glutamicum*  
US-09-738-626-5676

Query Match 41.1%; Score 37; DB 9; Length 1004;  
Best Local Similarity 58.3%; Pred. No. 5.8e+02;  
Matches 7; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

Qy 3 YGRDDNDPSPR 14  
Db 728 YGRDVEANGPSR 739

RESULT 35  
US-09-982-091A-2  
; Sequence 2, Application US/09982091A  
; Patent No. US20020151030A1  
; GENERAL INFORMATION:  
; APPLICANT: CALIFORNIA INSTITUTE OF TECHNOLOGY  
; APPLICANT: KUMAGAI, Akiko  
; APPLICANT: DUNPHY, William  
; TITLE OF INVENTION: CLASPIN PROTEINS AND METHODS OF USE THEREOF  
; FILE REFERENCE: CITI320-1  
; CURRENT APPLICATION NUMBER: US/09/982,091A  
; CURRENT FILING DATE: 2002-10-17  
; PRIOR APPLICATION NUMBER: US 60/241,246  
; PRIOR FILING DATE: 2000-10-17  
; NUMBER OF SEQ ID NOS: 12  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 2  
; LENGTH: 1285  
; TYPE: PRT  
; ORGANISM: *Xenopus laevis*  
US-09-982-091A-2

Query Match 41.1%; Score 37; DB 10; Length 1285;  
Best Local Similarity 63.6%; Pred. No. 7.5e+02;  
Matches 7; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 1 VGYGRDDNDP 11  
Db 641 VGDAEDNDDED 651

RESULT 36  
US-10-025-380-1068  
; Sequence 1068, Application US/10025380  
; Publication No. US20020182191A1  
; GENERAL INFORMATION:  
; APPLICANT: Xu, Jiangchun  
; APPLICANT: Lodes, Michael J.  
; APPLICANT: Secrist, Heather  
; APPLICANT: Benson, Darin R.  
; APPLICANT: Meagher, Madeleine Joy  
; APPLICANT: Stolk, John A.  
; APPLICANT: Wang, Tongtong  
; APPLICANT: Jiang, Yudi  
; APPLICANT: Smith, Carole L.  
; APPLICANT: King, Gordon E.  
; APPLICANT: Wang, Aljun  
; APPLICANT: Clapper, Jonathan D.  
; APPLICANT: Skeiky, Yasir A. W.  
; APPLICANT: Fanger, Gary R.  
; APPLICANT: Vedvick, Thomas S.  
; APPLICANT: Carter, Darick  
; TITLE OF INVENTION: COMPOUNDS FOR IMMUNOTHERAPY AND DIAGNOSIS  
; OF COLON CANCER AND METHODS FOR THEIR USE

```
; FILE REFERENCE: 210121.471C14
; CURRENT APPLICATION NUMBER: US/10/025.380
; CURRENT FILING DATE: 2001-12-19
; NUMBER OF SEQ ID NOS: 1129
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1068
; LENGTH: 5179
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-025-380-1068

Query Match      41.1%; Score 37; DB 9; Length 5179;
Best Local Similarity 66.7%; Pred. No. 3.3e+03;
Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY      4 GRDNDRDP 12      |||:|
Db      1317 GSDGDREP 1325

RESULT 37
US-09-922-217-1068
; Sequence 1068, Application US/09922217
; Patent No. US20020076414A1
; GENERAL INFORMATION:
; APPLICANT: Xu, Jiangchun
; APPLICANT: Lodes, Michael J.
; APPLICANT: Secrist, Heather
; APPLICANT: Benson, Darin R.
; APPLICANT: Meagher, Madeleine Joy
; APPLICANT: Stoik, John A.
; APPLICANT: Wang, Tongtong
; APPLICANT: Jiang, Yugu
; APPLICANT: Smith, Carole Lynn
; APPLICANT: King, Gordon E.
; APPLICANT: Wang, Aijun
; APPLICANT: Clapper, Jonathan D.
; TITLE OF INVENTION: COMPOUNDS FOR IMMUNOTHERAPY AND DIAGNOSIS
; TITLE OF INVENTION: OF COLON CANCER AND METHODS FOR THEIR USE
; FILE REFERENCE: 210121.471C13
; CURRENT APPLICATION NUMBER: US/09/922.217
; CURRENT FILING DATE: 2001-08-03
; NUMBER OF SEQ ID NOS: 1124
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1068
; LENGTH: 5179
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-922-217-1068

Query Match      41.1%; Score 37; DB 10; Length 5179;
Best Local Similarity 66.7%; Pred. No. 3.3e+03;
Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY      4 GRDNDRDP 12      |||:|
Db      1317 GSDGDREP 1325

RESULT 38
US-09-833-263-1068
; Sequence 1068, Application US/09833263
; Patent No. US20020110547A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Aijun
; APPLICANT: Clapper, Jonathan D.
; APPLICANT: Stoik, John A.
; APPLICANT: Meagher, Madeleine J.
; TITLE OF INVENTION: COMPOUNDS FOR IMMUNOTHERAPY AND
; TITLE OF INVENTION: DIAGNOSIS OF COLON CANCER AND METHODS FOR THEIR USE
; FILE REFERENCE: 210121.471C12
; CURRENT APPLICATION NUMBER: US/09/833.263
; CURRENT FILING DATE: 2001-04-10
```

```
; NUMBER OF SEQ ID NOS: 1093
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 1068
; LENGTH: 5179
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-833-263-1068

Query Match      41.1%; Score 37; DB 10; Length 5179;
Best Local Similarity 66.7%; Pred. No. 3.3e+03;
Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY      4 GRDNDRDP 12      |||:|
Db      1317 GSDGDREP 1325

RESULT 39
US-09-864-761-47955
; Sequence 47955, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharon G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
; TITLE OF INVENTION: GENE EXPRESSION ANALYSIS BY MICROARRAY
; FILE REFERENCE: Aecm1ca-x-1
; CURRENT APPLICATION NUMBER: US/09/864.761
; CURRENT FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/632,366
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: GB 24263.6
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US01/00666
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00664
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00669
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00665
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00668
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00663
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00662
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00661
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00670
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: US 09/608,408
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/774,203
; PRIOR FILING DATE: 2001-01-29
; NUMBER OF SEQ ID NOS: 49117
; SOFTWARE: Anomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 47955
; LENGTH: 69
; TYPE: PRT
; ORGANISM: Homo sapiens
```



```

; FEATURE:
; OTHER INFORMATION: MAP TO AC009962.1
; OTHER INFORMATION: EXPRESSED IN HBL100, SIGNAL = 1.1
; OTHER INFORMATION: EXPRESSED IN HE1A, SIGNAL = 1.3
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 2.2
; OTHER INFORMATION: EST HUMAN HIT: BE083092.1, EVALU 3.00e-24
; OTHER INFORMATION: SWISSPROT HIT: Q03638, EVALU 9.70e+00
US-09-864-761-47955

```

```

Query Match          40.6%; Score 36.5; DB 10; Length 69;
Best Local Similarity 47.1%; Pred. No. 39;
Matches 8; Conservative 4; Mismatches 2; Indels 3; Gaps 1;

```

```

Qy 1 VGYGRDDND--RDPSR 14
   :|:|:|:|:|:|:|:|:|
Db 19 LGFGRDEPDIAKIPSR 35

```

```

RESULT 40
US-09-741-669-405
; Sequence 405, Application US/09741669
; Patent No. US20020022718A1
; GENERAL INFORMATION:
; APPLICANT: Forsyth, R. Allyn
; APPLICANT: Ohlsen, Karl L.
; APPLICANT: Zyskind, Judith W.
; TITLE OF INVENTION: Genes identified as required for
; TITLE OF INVENTION: proliferation of E. coli
; FILE REFERENCE: ELITRA.009A
; CURRENT APPLICATION NUMBER: US/09/741,669
; CURRENT FILING DATE: 2000-12-19
; PRIOR APPLICATION NUMBER: US 60/173005
; PRIOR FILING DATE: 1999-12-23
; NUMBER OF SEQ ID NOS: 481
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 405
; LENGTH: 447
; TYPE: PRT
; ORGANISM: Escherichia coli
US-09-741-669-405

```

```

Query Match          40.6%; Score 36.5; DB 10; Length 447;
Best Local Similarity 61.5%; Pred. No. 2.9e+02;
Matches 8; Conservative 2; Mismatches 2; Indels 1; Gaps 1;

```

```

Qy 3 YGRDD-NDRDPSPR 14
   |||:|:|:|:|:|:|
Db 196 YGREGNDTRPSR 208

```

Search completed: April 9, 2003, 13:10:16  
 Job time : 8.90164 secs

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.4 p5.4578  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 9, 2003, 12:52:47 ; Search time 6.95082 Seconds  
(without alignments)  
221.291 Million cell updates/sec

Title: US-10-020-441-3

Perfect score: 90  
Sequence: 1 VGYGRDNDPRSRKN 16

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : PIR\_73:\*  
1: p1r1:\*  
2: p1r2:\*  
3: p1r3:\*  
4: p1r4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	90	100.0	264	2 A28942	pancreatic elastase
2	48.5	53.9	462	2 S54572	hypothetical prote
3	47	52.2	144	2 S69345	ovaduct-specific p
4	46	51.1	1428	1 ISBY12	DNA topoisomerase
5	45.5	50.6	422	2 T33186	hypothetical prote
6	45	50.0	810	2 T44430	protein PV100 [imp
7	44	48.9	443	2 C86345	hypothetical prote
8	44	48.9	561	2 T16148	hypothetical prote
9	44	48.9	8243	2 T11307	type I fatty acid
10	43	47.8	536	2 AH3227	geneC protein [imp
11	43	47.8	686	2 B87490	NADH dehydrogenase
12	42	46.7	576	2 B86316	protein T10022.11
13	42	46.7	655	2 A57681	hypothetical prote
14	42	46.7	995	2 T22942	hypothetical prote
15	42	46.7	1179	2 T04488	DNA topoisomerase
16	42	46.7	2559	2 T30850	fat facies protein
17	41	45.6	303	2 S28147	UI snRNP 70K prote
18	41	45.6	330	2 T34972	transcription repr
19	41	45.6	339	2 T11751	transcription repr
20	41	45.6	418	2 G87469	ThiJ/FfpI family p
21	41	45.6	427	2 S71367	UI snRNP 70K prote
22	41	45.6	745	2 C70848	probable lipoxygen
23	41	45.6	926	2 B96749	probable membrane
24	41	45.6	1245	2 S51255	translation initia
25	40.5	45.0	829	2 B64114	hypothetical prote
26	40	44.4	169	2 T33653	Acort-45 protein -
27	40	44.4	192	2 B72855	transporter BME120
28	40	44.4	216	2 AG3508	conserved hypothet
29	40	44.4	292	2 AF3032	

30	40	44.4	303	2 F98253	hypothetical prote
31	40	44.4	395	2 A54949	syndecan precursor
32	40	44.4	626	2 AG1256	DNA primase [impor
33	40	44.4	658	2 T40107	hypothetical 57.9
34	40	44.4	761	2 T24230	hypothetical prote
35	40	44.4	1057	2 H83273	ribonuclease E PA2
36	40	44.4	1271	2 A45555	glutamate rich pro
37	40	44.4	1306	2 S25370	MSB2 protein - yea
38	40	44.4	1325	2 T25753	hypothetical prote
39	39.5	43.9	532	2 G83424	hypothetical prote
40	39	43.3	112	2 S51129	genome polypeptin
41	39	43.3	135	1 CBRT5M	cytochrome B5, out
42	39	43.3	157	2 E75530	hypothetical prote
43	39	43.3	206	2 B64943	probable membrane
44	39	43.3	206	2 C90945	hypothetical prote
45	39	43.3	206	2 F85793	hypothetical prote

## ALIGNMENTS

## RESULT 1

A28942 pancreatic elastase (EC 3.4.21.36) precursor - fluke (Schistosoma mansoni)  
C/Species: Schistosoma mansoni  
C/Date: 30-Jun-1989 #sequence\_revision 30-Jun-1989 #text\_change 26-Aug-1999  
C/Accession: A28942  
R/Newport, G.R.; McKerrow, J.H.; Hedstrom, R.; Pettit, M.; McGarrigle, L.; Barr, P.J.;  
J. Biol. Chem. 263, 13179-13184, 1988  
A/Title: Cloning of the proteinase that facilitates infection by schistosome parasites.  
A/Reference number: A28942; MUID:8830818; PMID:3166457  
A/Accession: A28942  
A/Molecule type: mRNA  
A/Residues: 1-264 <NEM>  
A/Cross-references: GB:J03946; NID:G160947; PID:G160948  
C/Superfamily: trypsin; trypsin homology  
C/Keywords: hydrolase; serine proteinase  
F/27-256/Domain: trypsin homology <TRY>

Query Match 100.0%; Score 90; DB 2; Length 264;  
Best Local Similarity 100.0%; Pred. No. 2.3e-07;  
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 VGYGRDNDPRSRKN 16

Db 163 VGYGRDNDPRSRKN 178

## RESULT 2

S54572 hypothetical protein YMR11C - yeast (Saccharomyces cerevisiae)  
N/Alternate names: hypothetical protein YMR118.10c  
C/Species: Saccharomyces cerevisiae  
C/Date: 08-Jul-1995 #sequence\_revision 19-Oct-1995 #text\_change 19-Apr-2002  
C/Accession: S54572  
R/Hunt, S.; Bowman, S.  
submitted to the EMBL Data Library, May 1995  
A/Reference number: S54510  
A/Accession: S54572  
A/Molecule type: DNA  
A/Residues: 1-462 <HUN>  
A/Cross-references: EMBL:Z49702; NID:G817859; PIDN:CAA89747.1; PID:G817869; GSPDB:GN000  
A/Experimental source: strain AB972  
C/Genetics:  
A/Gene: MIPS:YMR11C  
A/Cross-references: SGD:S0004717  
A/Map position: 13R  
C/Superfamily: Saccharomyces cerevisiae hypothetical protein YMR11C  
Query Match 53.9%; Score 48.5; DB 2; Length 462;  
Best Local Similarity 55.6%; Pred. No. 3.2;  
Matches 10; Conservative 2; Mismatches 3; Indels 3; Gaps 1;

Oy 2 GYGR---DDNDRDPSRKN 16  
||| ||| ||| ||| |||  
Db 14 GYGSLDDDDSDRDSERRN 31

## RESULT 3

S69345  
cviduct-specific protein-1B - African clawed frog (fragment)  
C/Species: Xenopus laevis (African clawed frog)  
C/Date: 28-Oct-1996 #sequence\_revision 01-Dec-2000 #text\_change 01-Dec-2000  
C/Accession: S69345  
R/Mehta, R.J.; Tata, J.R.  
Eur. J. Biochem. 229, 224-232, 1995  
A/Title: Structural and functional characterization and cloning of Xenopus FOSP-1 (frog  
A/Reference number: S69342; MUID:95262671; PMID:7744034  
A/Accession: S69343  
A/Status: preliminary  
A/Molecule type: mRNA  
A/Residues: 1-144 <MEH>  
A/Cross-references: EMBL:X81570; NID:G587595; PIDN:CA57266.1; PID:G587596  
A/Accession: S69345  
A/Status: preliminary; translation not shown  
A/Molecule type: DNA  
A/Residues: 1-24 <ME2>  
A/Cross-references: EMBL:X81572; NID:G587597; PIDN:CA57268.1; PID:G587598

Query Match 52.2%; Score 47; DB 2; Length 144;  
Best Local Similarity 53.3%; Pred. No. 1.6;  
Matches 8; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

Oy 2 GYGRDNDPRDPSRKN 16  
||| ||| ||| ||| |||  
Db 91 GIGRKDNKNGDKKN 105

## RESULT 4

ISBT72  
DNA topoisomerase (ATP-hydrolyzing) (BC 5.99.1.3) - yeast (Saccharomyces cerevisiae)  
N/Alternate names: DNA gyrase; DNA topoisomerase II; protein YNL088W  
C/Species: Saccharomyces cerevisiae  
C/Date: 30-Sep-1992 #sequence\_revision 10-Nov-1995 #text\_change 16-Jun-2000  
C/Accession: S57534; A25630; S63027; S30866; S65093  
R/Soler-Mira, A.; Saiz, J.E.; Ballesta, J.P.G.; Remacha, M.  
submitted to the EMBL Data Library, June 1995  
A/Reference number: S57533  
A/Accession: S57534  
A/Molecule type: DNA  
A/Residues: 1-1428 <SOL>  
A/Cross-references: EMBL:X89016; NID:G887621; PIDN:CA61422.1; PID:G887623  
R/Glaever, G.; Lynn, R.; Goto, T.; Wang, J.C.  
J. Biol. Chem. 261, 12448-12454, 1986  
A/Title: The complete nucleotide sequence of the structural gene TOP2 of yeast DNA topoi  
A/Reference number: A25630; MUID:86304413; PMID:3017975  
A/Accession: A25630  
A/Molecule type: DNA  
A/Residues: 1-74, 'N', '75-546, 'L', '548-836, 'R', '838-1428 <GIA>  
A/Cross-references: GB:M13814; NID:G172997; PIDN:AAB3610.1; PID:G172998  
R/Soler-Mira, A.; Saiz, J.E.; Ballesta, J.P.G.; Remacha, M.  
submitted to the protein Sequence Database, April 1996  
A/Reference number: S63018  
A/Accession: S63027  
A/Molecule type: DNA  
A/Residues: 1-1428 <SOW>  
A/Cross-references: EMBL:Z71364; NID:G1301988; PIDN:CA95964.1; PID:G1301989; GSPDB:GN00  
R/Jamatiipour, M.; Liu, Y.X.; Nittis, J.L.  
submitted to the EMBL Data Library, January 1993  
A/Description: The top2-5 mutant of yeast topoisomerase II encodes an enzyme resistant t  
A/Reference number: S30866  
A/Accession: S30866  
A/Molecule type: mRNA  
A/Residues: 812-836, 'R', '838-882, 'P', '884, 'IT', '887-977 <JAN>  
A/Cross-references: EMBL:L08968; NID:G172999; PIDN:AAB59328.1; PID:G173000

R/Soler-Mira, A.; Saiz, J.E.; Ballesta, J.P.G.; Remacha, M.  
Yeast 12, 485-491, 1996  
A/Title: The sequence of a 17 933 bp segment of Saccharomyces cerevisiae chromosome XIV  
A/Reference number: S65092; MUID:96310628; PMID:8740422  
A/Accession: S65093  
A/Status: nucleic acid sequence not shown  
A/Molecule type: DNA  
A/Residues: 1-1428 <SOF>  
A/Cross-references: EMBL:X89016; NID:G887621; PIDN:CA61422.1; PID:G887623  
C/Comment: Type II DNA topoisomerase catalyzes the ATP-dependent transient breakage, pa  
C/Keywords: ATP; DNA binding; DNA replication; heterotetramer; isomerase; phosphoprotein.

Query Match 51.1%; Score 46; DB 1; Length 1428;  
Best Local Similarity 72.7%; Pred. No. 28;  
Matches 8; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Oy 6 DDNDRDPSRKN 16  
||| ||| ||| ||| |||  
Db 1198 DDEYDPSRKN 1208

## RESULT 5

T33186  
hypochemical protein F22F7.2 - Caenorhabditis elegans  
C/Species: Caenorhabditis elegans  
C/Date: 29-Oct-1999 #sequence\_revision 29-Oct-1999 #text\_change 02-Sep-2000  
C/Accession: T33186  
R/Miller, N.; Kramer, J.; Smith, A.  
submitted to the EMBL Data Library, May 1998  
A/Description: The sequence of C. elegans cosmid F22F7.  
A/Reference number: Z21299  
A/Accession: T33186  
A/Status: preliminary; translated from GB/EMBL/DBJ  
A/Molecule type: DNA  
A/Residues: 1-422 <MLL>  
A/Cross-references: EMBL:AF067937; PIDN:AA19222.1; GSPDB:GN00023; CESP:F22F7.2  
A/Experimental source: Strain Bristol N2; clone F22F7  
C/Genetics:  
A/Gene: CESP:F22F7.2  
A/Map position: 5  
A/Intons: 53/3; 93/3; 180/3; 316/3; 388/2  
C/Superfamily: Streptomyces coelicolor conserved hypochemical protein scSFZA.12c

Query Match 50.6%; Score 45.5; DB 2; Length 422;  
Best Local Similarity 60.0%; Pred. No. 9.2;  
Matches 9; Conservative 2; Mismatches 3; Indels 1; Gaps 1;

Oy 2 GYGR-DDNDRDPSRKN 15  
||| ||| ||| ||| |||  
Db 338 GYGRDPRDRPKKK 352

## RESULT 6

T44430  
protein PVI00 (imported) - winter squash  
C/Species: Cucurbita maxima (winter squash)  
C/Date: 21-Jan-2000 #sequence\_revision 21-Jan-2000 #text\_change 21-Jul-2000  
C/Accession: T44430  
R/Yanada, K.; Shinada, T.; Kondo, M.; Nishimura, M.; Hara-Nishimura, I.  
J. Biol. Chem. 274, 2563-2570, 1999  
A/Title: Multiple functional proteins are produced by cleaving Asn-Gln bonds of a singl  
A/Reference number: Z22767; MUID:99107919; PMID:9891029  
A/Accession: T44430  
A/Status: preliminary; translated from GB/EMBL/DBJ  
A/Molecule type: mRNA  
A/Residues: 1-810 <YAM>  
A/Cross-references: EMBL:AB019195; NID:G3808061; PIDN:BAA34056.1; PID:G3808062

Query Match 50.0%; Score 45; DB 2; Length 810;  
Best Local Similarity 63.6%; Pred. No. 22;  
Matches 7; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 5 RDDNDRPSRK 15  
|||:||||:  
DB 158 RDENERDPKRE 168

## RESULT 7

C86345  
hypothetical protein F16F4.11 - Arabidopsis thaliana

C/Species: Arabidopsis thaliana (mouse-ear cress)

C/Date: 02-Mar-2001 #sequence\_revision 02-Mar-2001 #text\_change 31-Dec-2001

C/Accession: C86345

R/Theologis, A.; Ecker, J.R.; Palm, C.J.; Federpiel, N.A.; Kaul, S.; White, O.; Alonso, Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, K.; ansen, N.F.; Hughes, B.; Hutzar, L.

A/Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Maiti, R.; Marziani, Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.; Shim, P.; Southwick, A.M.; Sun, H.; Tallon, ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.

A/Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.

A/Reference number: A86141; MUID:21016719; PMID:11130712

A/Accession: C86345

A/Status: preliminary

A/Molecule type: DNA

A/Residues: 1-443 <STO>

A/Cross-references: GB:AE05172; NID:g8920640; PIDN:AF81362.1; GSPDB:GN00141

C/Genetics:

A/Map position: 1

Query Match 48.9%; Score 44; DB 2; Length 443;  
Best Local Similarity 66.7%; Pred. No. 17;  
Matches 8; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 5 RDDNDRPSRK 16  
|||:||||:  
DB 265 RDDNDRPSRK 276

## RESULT 8

T16148  
hypothetical protein F25B5.7 - Caenorhabditis elegans

C/Species: Caenorhabditis elegans

C/Date: 20-Sep-1999 #sequence\_revision 20-Sep-1999 #text\_change 20-Sep-1999

C/Accession: T16148

R/Talch, A.

submitted to the EMBL Data Library, March 1995

A/Description: The sequence of C. elegans cosmid F25B5.

A/Reference number: Z18468

A/Accession: T16148

A/Status: preliminary; translated from GB/EMBL/DBJ

A/Molecule type: DNA

A/Residues: 1-561 <TAI>

A/Cross-references: EMBL:U23172; NID:g726388; PID:g726394; PIDN:AAC46528.1; CESP:F25B5.7

A/Experimental source: strain Bristol N2

C/Genetics:

A/Gene: CESP:F25B5.7

A/Intons: 39/2; 157/3; 258/2; 290/2; 356/1; 531/1

Query Match 48.9%; Score 44; DB 2; Length 561;  
Best Local Similarity 47.6%; Pred. No. 22;  
Matches 10; Conservative 1; Mismatches 2; Indels 8; Gaps 1;

QY 4 GRDD-----NDRPSRK 16  
|||:||||:  
DB 13 GRDDSSRVAKAENDRKPGRN 33

## RESULT 9

T31307  
type I fatty acid synthase homolog - Cryptosporidium parvum

C/Species: Cryptosporidium parvum

C/Date: 22-Oct-1999 #sequence\_revision 22-Oct-1999 #text\_change 17-Nov-2000

C/Accession: T31307

R/Zhu, G.; Marchewka, M.J.; Woods, K.M.; Upton, S.J.; Keithly, J.S.

submitted to the EMBL Data Library, August 1998

A/Description: Characterization of a type I FAS gene in the parasitic protozoan Cryptos

A/Reference number: Z20993

A/Accession: T31307

A/Status: preliminary; translated from GB/EMBL/DBJ

A/Molecule type: DNA

A/Residues: 1-8243 <ZHU>

A/Cross-references: EMBL:AF082993; NID:g4092068; PID:g4092069; PIDN:AAC99407.1

C/Genetics:

A/Note: FAS1

C/Superfamily: 3-oxoacyl-[acyl-carrier-protein] synthase I homology; acetate-CoA ligase

C/Keywords: carrier protein

F/100-714/Domain: acyl carrier protein homology <ACPL>

F/791-861/Domain: acyl carrier protein homology <ACPL>

F/906-1308/Domain: 3-oxoacyl-[acyl-carrier-protein] synthase I homology <OAS1>

F/2924-2992/Domain: acyl carrier protein homology <ACR2>

F/3062-3463/Domain: 3-oxoacyl-[acyl-carrier-protein] synthase I homology <OAS2>

F/5157-5227/Domain: acyl carrier protein homology <ACP3>

F/5274-5689/Domain: 3-oxoacyl-[acyl-carrier-protein] synthase I homology <OAS3>

Query Match 48.9%; Score 44; DB 2; Length 8243;  
Best Local Similarity 57.1%; Pred. No. 4e+02;  
Matches 8; Conservative 1; Mismatches 5; Indels 0; Gaps 0;

QY 3 YGRDNDRPSRK 16  
|||:||||:  
DB 4317 YGRDQELPSPSN 4330

## RESULT 10

AH3227  
gene protein [imported] - Agrobacterium tumefaciens (strain C58, Dupont) plasmid Ti

C/Species: Agrobacterium tumefaciens

C/Date: 11-Jan-2002 #sequence\_revision 11-Jan-2002 #text\_change 11-Jan-2002

C/Accession: AH3227

R/Wood, D.W.; Setubal, J.C.; Kaul, R.; Monks, D.; Chen, L.; Wood, G.E.; Chen, Y.; Moo, erage, G.; Gillet, W.; Grant, C.; Guenther, D.; Kuttyavin, T.; Levy, R.; Li, M.; McClel

; Karp, P.; Romero, P.; Zhang, S.

Science 294, 2317-2323, 2001

A/Authors: Yoo, H.; Tao, Y.; Biddle, P.; Jung, M.; Krespan, W.; Perry, M.; Gordon-Kamm, ster, E.W.

A/Title: The Genome of the Natural Genetic Engineer Agrobacterium tumefaciens C58.

A/Reference number: AB2577; PMID:11743193

A/Accession: AH3227

A/Status: preliminary

A/Molecule type: DNA

A/Residues: 1-536 <KUR>

A/Cross-references: GB:AB08690; PIDN:PAL46238.1; PID:g17744015; GSPDB:GN00189

A/Experimental source: strain C58 (Dupont)

C/Genetics:

A/Gene: geneC

A/Genome: plasmid

Query Match 47.8%; Score 43; DB 2; Length 536;  
Best Local Similarity 66.7%; Pred. No. 31;  
Matches 8; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 1 VGYGRDNDRDP 12  
|||:||||:  
DB 391 VGYRADNDNDRSP 402

## RESULT 11

E87490  
NADH dehydrogenase I, G subunit C01946 [imported] - Caulobacter crescentus  
C/Species: Caulobacter crescentus  
C/Date: 20-Apr-2001 #sequence\_revision 20-Apr-2001 #text\_change 10-May-2001



RESULT 16  
T30850  
fat facets protein homolog Fam - mouse  
C/Species: Mus musculus (house mouse)  
C/Date: 22-Oct-1999 #sequence\_revision 22-Oct-1999 #text\_change 22-Oct-1999  
C/Accession: T30850  
R/Wood, S.A.  
submitted to the EMBL Data Library, August 1996  
A/Description: Fam, fat facets in mouse.  
A/Reference number: Z20902  
A/Accession: T30850  
A/Status: preliminary; translated from GB/EMBL/DBJ  
A/Molecule type: mRNA  
A/Residues: 1-2559 <MO>  
A/Cross-references: EMBL:U67874; NID:G1527198; PID:G1527199; PIDN:AA07731.1  
C/Genetics:  
A/Gene: Fam  
A/Map position: X

Query Match 46.7%; Score 42; DB 2; Length 2559;  
Best Local Similarity 70.0%; Pred. No. 2.4e+02;  
Matches 7; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

OY 3 YGRDNDNDP 12  
DB 630 YGRDNDNDP 639

RESULT 17  
S28147  
UI snRNP 70K protein - Arabidopsis thaliana (fragment)  
N/Alternate names: UI small nuclear ribonucleoprotein 70K chain  
C/Species: Arabidopsis thaliana (mouse-ear cress)  
C/Date: 20-Feb-1995 #sequence\_revision 01-Nov-1996 #text\_change 07-May-1999  
C/Accession: S28147  
R/Reddy, A.S.N.; Czernik, A.J.; An, G.; Poovaiah, B.W.  
Biochim. Biophys. Acta 1171, 88-92, 1992  
A/Title: Cloning of the cDNA for UI small nuclear ribonucleoprotein particle 70K protein  
A/Reference number: S28147; MUID:93042012; PMID:1420366  
A/Accession: S28147  
A/Status: not compared with conceptual translation  
A/Molecule type: mRNA  
A/Residues: 1-303 <RD>  
C/Superfamily: unassigned ribonucleoprotein repeat-containing proteins; ribonucleoproteins  
F/33-100/domain: ribonucleoprotein repeat homology <RR>

Query Match 45.6%; Score 41; DB 2; Length 303;  
Best Local Similarity 62.5%; Pred. No. 36;  
Matches 10; Conservative 1; Mismatches 1; Indels 4; Gaps 1;

OY 4 GRD---DNDPDRSRK 15  
DB 228 GRDTRSDHDDRSRK 243

RESULT 18  
T34972  
probable membrane protein - Streptomyces coelicolor  
C/Species: Streptomyces coelicolor  
C/Date: 05-Nov-1999 #sequence\_revision 05-Nov-1999 #text\_change 05-Nov-1999  
C/Accession: T34972  
R/Sundera, D.C.; Harris, D.; James, K.D.; Parkhill, J.; Barrell, B.G.; Rajandream, M.A.  
submitted to the EMBL Data Library, August 1999  
A/Reference number: Z21563  
A/Accession: T34972  
A/Status: preliminary; translated from GB/EMBL/DBJ  
A/Molecule type: DNA  
A/Residues: 1-330 <SAU>  
A/Cross-references: EMBL:AJ09663; PIDN:CA852011.1; GSPDB:GN00070; SCOEDB:SCA410.35C  
A/Experimental source: strain AJ(2)  
C/Genetics:  
A/Gene: SCOEDB:SCA410.35c

Query Match 45.6%; Score 41; DB 2; Length 330;  
Best Local Similarity 56.2%; Pred. No. 39;  
Matches 9; Conservative 1; Mismatches 2; Indels 4; Gaps 1;

OY 3 YGRDNDNDP---PSR 14  
DB 135 YGRDNDNDPVTTPAR 150

RESULT 19  
T11751  
transcription repressor RRM1 - kidney bean  
N/Alternate names: G-box binding factor  
C/Species: Phaseolus vulgaris (kidney bean)  
C/Date: 16-Jul-1999 #sequence\_revision 16-Jul-1999 #text\_change 18-Aug-2000  
C/Accession: T11751  
R/Chern, M.S.; Eiben, H.G.; Busto, M.M.  
Plant J. 10, 135-148, 1996  
A/Title: The developmentally regulated bZIP factor RRM1 modulates transcription from le  
A/Reference number: Z17333; MUID:96314865; PMID:8758983  
A/Accession: T11751  
A/Status: preliminary; translated from GB/EMBL/DBJ  
A/Molecule type: mRNA  
A/Residues: 1-339 <CHE>  
A/Cross-references: EMBL:U57389; NID:G1354856; PIDN:AA36514.1; PID:G1354857  
C/Function:  
A/Description: modulates transcription from lectin and storage protein genes  
C/Superfamily: fava bean G-box-binding protein; fos/jun DNA-binding domain homology  
C/Keywords: DNA binding; leucine zipper; transcription regulation

Query Match 45.6%; Score 41; DB 2; Length 339;  
Best Local Similarity 57.1%; Pred. No. 40;  
Matches 8; Conservative 1; Mismatches 5; Indels 0; Gaps 0;

OY 2 YGRDNDNDPDRSRK 15  
DB 109 YGRDNDNDPDRSRK 122

RESULT 20  
G87469  
Thi/PfPI family protein [imported] - Caulobacter crescentus  
C/Species: Caulobacter crescentus  
C/Date: 20-Apr-2001 #sequence\_revision 20-Apr-2001 #text\_change 20-Apr-2001  
C/Accession: G87469  
R/Nieman, W.C.; Feldblyum, T.V.; Paulsen, I.T.; Nelson, K.E.; Eisen, J.; Heidelberg, J.  
B.; Laub, M.T.; Deboy, R.T.; Dodson, R.J.; Durkin, A.S.; Gwinn, M.U.; Hatt, D.H.; Kolo  
n, J.; Ermolaeva, M.; White, O.; Salzberg, S.L.; Shapiro, L.; Venter, J.C.; Fraser, C.M.  
Proc. Natl. Acad. Sci. U.S.A. 98, 4136-4141, 2001  
A/Title: Complete Genome Sequence of Caulobacter crescentus.  
A/Reference number: A87249; MUID:21173698; PMID:11259647  
A/Accession: G87469  
A/Status: preliminary  
A/Molecule type: DNA  
A/Residues: 1-418 <STO>  
A/Cross-references: GB:AE005673; NID:G13423207; PIDN:AAK3755.1; GSPDB:GN00148  
C/Genetics:  
A/Gene: CCI1779

Query Match 45.6%; Score 41; DB 2; Length 418;  
Best Local Similarity 72.7%; Pred. No. 51;  
Matches 8; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 5 RDDNDNDPDRSRK 15  
DB 407 RDDNDNDPDRSRK 417

RESULT 21  
S71367  
UI snRNP 70K protein - Arabidopsis thaliana  
N/Alternate names: UI small nuclear ribonucleoprotein 70K chain  
C/Species: Arabidopsis thaliana (mouse-ear cress)

C>Date: 28-Oct-1996 #sequence\_revision 27-Feb-1997 #text\_change 18-Feb-2000  
 C/Accession: S71367; T46144; S71411  
 R/Golovkin, M.; Reddy, A.S.N.  
 submitted to the EMBL Data Library, March 1996  
 A/Reference number: S71367  
 A/Accession: S71367  
 A/Molecule type: mRNA  
 A/Residues: 1-427 <COL>  
 A/Cross-references: EMBL:M93439; NID:g1255710; PIDN:AD12773.1; PID:g1255711  
 R/Bloeker, H.; Mewes, H.W.; Mayer, K.F.X.; Lemcke, K.; Schuelter, C.; Quetier, F.; Sala  
 submitted to the Protein Sequence Database, December 1999  
 A/Reference number: Z23024  
 A/Accession: T46144  
 A/Status: preliminary  
 A/Molecule type: DNA  
 A/Residues: 1-427 <BLO>  
 A/Cross-references: EMBL:AL132979  
 A/Experimental source: cultivar Columbia; BAC clone T3A5  
 C/Genetics:  
 A/Gene: T3A5.50  
 A/Map position: 3  
 A/Insertion: 33/3; 72/1; 98/3; 124/1; 145/3; 166/3; 194/1; 265/2  
 C/Superfamily: unassigned ribonucleoprotein repeat-containing proteins; ribonucleoprotein  
 F;139-206/Domain: ribonucleoprotein repeat homology <RRM>

Query Match 45.6%; Score 41; DB 2; Length 427;  
 Best Local Similarity 62.5%; Pred. No. 52;  
 Matches 10; Conservative 1; Mismatches 1; Indels 4; Gaps 1;

QY 4 GRD---DNDRDPGRK 15  
 ||| :||| |||  
 Db 335 GRDRTSRDHRDRGRK 350

RESULT 22  
 C70848  
 probable icd2 protein - Mycobacterium tuberculosis (strain H37RV)  
 C/Species: Mycobacterium tuberculosis  
 C/Date: 17-Jul-1998 #sequence\_revision 17-Jul-1998 #text\_change 20-Jun-2000  
 C/Accession: C70848  
 R/Cole, S.T.; Broesch, R.; Parthill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.  
 ; Connor, R.; Davies, R.; Devlin, K.; Felwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.  
 Rajendram, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.  
 Nature 393, 537-544, 1998  
 A/Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.  
 A/Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome  
 A/Reference number: A70500; MUID:98295987; PMID:9634230  
 A/Accession: C70848  
 A/Status: preliminary; nucleic acid sequence not shown; translation not shown  
 A/Molecule type: DNA  
 A/Residues: 1-745 <COL>  
 A/Cross-references: GB:AL021428; GB:AL123456; NID:g3261514; PIDN:CA16247.1; PID:g280872  
 C/Experimental source: strain H37RV  
 C/Genetics:  
 A/Gene: icd2  
 C/Superfamily: Vibrio isocitrate dehydrogenase (NADP+)

Query Match 45.6%; Score 41; DB 2; Length 745;  
 Best Local Similarity 77.8%; Pred. No. 94;  
 Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 7 DNDRDPGRK 15  
 ||| :||| |||  
 Db 645 DNDKSPSRK 653

RESULT 23  
 E96749  
 probable lipoxxygenase T10D10.1 [imported] - Arabidopsis thaliana  
 C/Species: Arabidopsis thaliana (mouse-ear cress)  
 C/Date: 02-Mar-2001 #sequence\_revision 02-Mar-2001 #text\_change 23-Mar-2001  
 C/Accession: E96749  
 R/Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso,

Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, K.  
 ansen, N.F.; Hughes, B.; Hutzar, L.  
 Nature 408, 816-820, 2000  
 A/Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.  
 C.A.; Li, J.H.; Li, Y.; Liu, X.; Liu, S.X.; Liu, Z.A.; Lueros, J.S.; Matti, R.; Marziani  
 Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.  
 A/Authors: Salzberg, S.L.; Schwartz, J.R.; Shim, P.; Southwick, A.M.; Sun, H.; Tallon,  
 ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.  
 A/Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.  
 A/Reference number: A86141; MUID:21016719; PMID:11130712  
 A/Accession: E96749  
 A/Status: preliminary  
 A/Molecule type: DNA  
 A/Residues: 1-926 <STO>  
 A/Cross-references: GB:AE005173; NID:g6730749; PIDN:AAF27138.1; GSPDB:GN00141  
 C/Genetics:  
 A/Gene: T10D10.1  
 A/Map position: 1  
 C/Superfamily: lipoxxygenase

Query Match 45.6%; Score 41; DB 2; Length 926;  
 Best Local Similarity 63.6%; Pred. No. 1,2e+02;  
 Matches 7; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 6 DNDRDPGRKN 16  
 ||| :||| |||  
 Db 885 DKNRDPGRKN 895

RESULT 24  
 S51255  
 probable membrane protein YDR104C - Yeast (Saccharomyces cerevisiae)  
 N/Alternate names: hypothetical protein YD8557.13c  
 C/Species: Saccharomyces cerevisiae  
 C/Date: 23-Feb-1995 #sequence\_revision 12-May-1995 #text\_change 19-Apr-2002  
 C/Accession: S51255  
 R/Murphy, L.; Harris, D.  
 submitted to the EMBL Data Library, January 1995  
 A/Reference number: S51243  
 A/Accession: S51255  
 A/Molecule type: DNA  
 A/Residues: 1-1245 <MUR>  
 A/Cross-references: EMBL:Z47746; NID:g633627; PID:g633640; GSPDB:GN00004; MIPS:YDR104C  
 C/Genetics:  
 A/Gene: SGD:SP071; MIPS:YDR104C  
 A/Cross-references: SGD:S0002511  
 A/Map position: 4R  
 C/Superfamily: Saccharomyces cerevisiae probable membrane protein YDR104C  
 C/Keywords: transmembrane protein  
 F;1045-1061/Domain: transmembrane #status predicted <TMM>

Query Match 45.6%; Score 41; DB 2; Length 1245;  
 Best Local Similarity 70.0%; Pred. No. 1.6e+02;  
 Matches 7; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 3 YGRDNDNDP 12  
 ||| :||| |||  
 Db 1165 YGRDNDNDP 1174

RESULT 25  
 E64114  
 translation initiation factor IF-2 - Haemophilus influenzae (strain Rd KW20)  
 C/Species: Haemophilus influenzae  
 C/Date: 18-Aug-1995 #sequence\_revision 18-Aug-1995 #text\_change 02-Feb-2001  
 C/Accession: E64114  
 R/Fleischmann, R.D.; Adams, M.D.; White, O.; Clayton, R.A.; Kirkness, E.F.; Kerlavage,  
 J.; Gokey, J.D.; Scott, J.; Shirley, R.; Liu, L.I.; Glodek, A.; Kelley, J.M.; Weidman,  
 D.M.; Brandon, R.C.; Fine, L.D.; Fitchman, J.L.; Fuhrmann, J.L.; Geoghagen, N.S.M.  
 Science 269, 496-512, 1995  
 A/Authors: Gnehm, C.L.; McDonald, L.A.; Small, K.V.; Frazer, C.M.; Smith, H.O.; Venter,  
 A/Title: Whole-genome random sequencing and assembly of Haemophilus influenzae Rd.  
 A/Reference number: A64000; MUID:95350630; PMID:7542800



```

A:Accession: E64114
A>Status: nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-829 <TIGR>
A:Cross-references: GB:U52808; GB:I42023; NID:G1574739; PID:AA02993.1; PID:G1574742; J
C:Species: Caenorhabditis elegans
C:Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 29-Oct-1999
C:Accession: T33653
R:Zhu, H.; Hinds, K.; Keppler, D.
submitted to the EMBL data library, October 1998
A:Description: The sequence of C. elegans cosmid T06A10.
A:Reference number: Z21382
A:Accession: T33653
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-169 <ZHU>
A:Cross-references: EMBL:AF09926; PID:AA068904.1; GSPDB:GN00022; CESP:T06A10.3
A:Experimental source: strain Bristol N2; clone T06A10
C:Genetic8:
A:Gene: CESP:T06A10.3
A:Map position: 4
A:Introns: 28/3; 57/2

Query Match          44.4%; Score 40; DB 2; Length 169;
Best Local Similarity 53.8%; Pred. No. 28;
Matches 7; Conservative 1; Mismatches 5; Indels 0; Gaps 0;

QY 4 GRDDNDPDSRKN 16
      |||||
Db 120 GSDSDDEDEERN 132

RESULT 27
E72855
ACORF-45 protein - Autographa californica nuclear polyhedrosis virus
C:Species: Autographa californica nuclear polyhedrosis virus, AcMNPV
A:Note: dsDNA virus
C:Date: 12-Nov-1999 #sequence_revision 12-Nov-1999 #text_change 12-Nov-1999
C:Accession: E72855
R:AYres, M.D.; Howard, S.C.; Kuzio, J.; Lopez-Ferber, M.; Possee, R.D.
Virology 202, 586-605, 1994
A:Title: The complete DNA sequence of Autographa californica nuclear polyhedrosis virus.
A:Reference number: A72850; MUID:94303173; PMID:8030224
A:Accession: E72855
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-192 <AYR>
A:Cross-references: GB:L22850; NID:G510708; PID:AAA66675.1; PID:G559114
C:Genetic8:
A:Gene: ACORF-45

Query Match          44.4%; Score 40; DB 2; Length 192;
Best Local Similarity 87.5%; Pred. No. 32;

```

```

Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 5 RDDNDRDP 12
|||||
db 50 RDDNDDDP 57

RESULT 28
AG3508
transporter BMEI2053 [imported] - Brucella melitensis (strain 16M)
C:Species: Brucella melitensis
C:Date: 01-Feb-2002 #sequence_revision 01-Feb-2002 #text_change 01-Feb-2002
C:Accession: AG3508
R:DeVechchio, V.G.; Kaparat, V.; Redkar, R.U.; Patra, G.; Mujer, C.; Los, T.; Ivanova,
.; Mazur, M.; Goltzman, E.; Selkov, E.; Elzer, P.H.; Hagins, S.; O'Callaghan, D.; Letesce
Proc. Natl. Acad. Sci. U.S.A. 99, 443-448, 2002
A:Title: The genome sequence of the facultative intracellular pathogen Brucella meliten
A:Reference number: AJ3252; PMID:11756688
A:Accession: AG3508
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-216 <KUR>
A:Cross-references: GB:AE008917; PIDN:AAL53234.1; PID:G17984111; GSPDB:GN00190
C:Genetics:
A:Gene: BMEI2053
A:Map position: 1

Query Match 44.4%; Score 40; DB 2; Length 216;
Best Local Similarity 53.8%; Pred. No. 36;
Matches 7; Conservative 2; Mismatches 4; Indels 0; Gaps 0;
QY 3 YGRDNDNRDP 15
|||||
db 196 FARDTNSRDP 208

RESULT 29
AF3032
conserved hypothetical protein Atu3867 [imported] - Agrobacterium tumefaciens (strain C
C:Species: Agrobacterium tumefaciens
C:Date: 11-Jan-2002 #sequence_revision 11-Jan-2002 #text_change 11-Jan-2002
C:Accession: AF3032
R:Wood, D.W.; Setubal, J.C.; Kaul, R.; Monks, D.; Chen, L.; Wood, G.E.; Chen, Y.; Woo,
erage, G.; Gillet, W.; Grant, C.; Guenther, D.; Kutayavin, T.; Levy, R.; Li, M.; McClell
; Karp, P.; Romero, P.; Zhang, S.
Science 294, 2317-2323, 2001
A:Authors: Yoo, H.; Tao, Y.; Biddle, P.; Jung, M.; Kreepan, W.; Perry, M.; Gordon-Kamm,
ber, E.W.
A:Title: The Genome of the Natural Genetic Engineer Agrobacterium tumefaciens C58.
A:Reference number: AB2577; PMID:11743193
A:Accession: AF3032
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-292 <KUR>
A:Cross-references: GB:AE008689; PIDN:AAL44676.1; PID:G17742303; GSPDB:GN00187
A:Experimental source: strain C58 (Dupont)
C:Genetics:
A:Gene: Atu3867
A:Map position: linear chromosome

Query Match 44.4%; Score 40; DB 2; Length 292;
Best Local Similarity 66.7%; Pred. No. 50;
Matches 8; Conservative 1; Mismatches 3; Indels 0; Gaps 0;
QY 1 VGYGRDNDNRDP 12
|||||
db 258 VGYGRDSDVAP 269

RESULT 30
198253
hypothetical protein AGR_L1957 [imported] - Agrobacterium tumefaciens (strain C58, Cer

```

C/Species: *Agrobacterium tumefaciens*  
 C/Date: 22-Oct-2001 #sequence\_revision 22-Oct-2001 #text\_change 11-Jan-2002  
 C/Accession: F98253  
 R/Goodner, B.; Hinkle, G.; Gatungu, S.; Miller, N.; Blanchard, M.; Quirollo, B.; Goldman, A.; Liu, F.; Wollam, C.; Allinger, M.; Doughty, D.; Scott, C.; Lappas, C.; Markelz, B.; Science 294, 2323-2328, 2001  
 A/Title: Genome Sequence of the Plant Pathogen and Biotechnology Agent *Agrobacterium tumefaciens*  
 A/Reference number: A97359; PMID:11743194  
 A/Accession: F98253  
 A/Status: preliminary  
 A/Molecule type: DNA  
 A/Residues: 1-303 <KIR>  
 A/Cross-references: GB:AE007870; PIDN:AA89552.1; PID:g15159434; GSPDB:GN00170  
 C/Genetics:  
 A/Gene: AGR L.1957  
 A/Map position: linear chromosome

Query Match 44.4%; Score 40; DB 2; Length 303;  
 Best Local Similarity 66.7%; Pred. No. 52;  
 Matches 8; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

OY 1 VGYGRDNDPRD 12  
 Db 269 VGYGRDNDPRD 280

RESULT 31  
 A54949  
 synedcan precursor - fruit fly (*Drosophila melanogaster*)  
 C/Species: *Drosophila melanogaster*  
 C/Date: 11-Nov-1994 #sequence\_revision 11-Nov-1994 #text\_change 21-Jul-2000  
 C/Accession: A54949  
 R/Spring, J.; Paine-Sanders, S.E.; Hynes, R.O.; Bernfield, M. Proc. Natl. Acad. Sci. U.S.A. 91, 3334-3338, 1994  
 A/Title: *Drosophila synedcan*: conservation of a cell-surface heparan sulfate proteoglycan  
 A/Reference number: A54949; MUID:94211858; PMID:8159748  
 A/Accession: A54949  
 A/Status: preliminary  
 A/Molecule type: mRNA; protein  
 A/Residues: 1-395 <SPR>  
 A/Cross-references: GB:U03282; NID:g437282; PIDN:AA034307.1; PID:g437283  
 A/Note: parts of this sequence were confirmed by protein sequencing  
 C/Genetics:  
 A/Gene: FlyBase:Sdc  
 A/Cross-references: FlyBase:FBgn010415  
 C/Keywords: transmembrane protein

Query Match 44.4%; Score 40; DB 2; Length 395;  
 Best Local Similarity 46.2%; Pred. No. 70;  
 Matches 6; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

OY 2 GYGRDNDPRD 14  
 Db 82 GYGRDNDPRD 94

RESULT 32  
 AG1256  
 DNA primase [imported] - *Listeria monocytogenes* (strain EGD-e)  
 C/Species: *Listeria monocytogenes*  
 C/Date: 27-Nov-2001 #sequence\_revision 27-Nov-2001 #text\_change 27-Nov-2001  
 C/Accession: AG1256  
 R/Glaser, P.; Frangeul, L.; Buchrieser, C.; Amend, A.; Baquero, F.; Berche, P.; Bioecker, J.; Dominguez-Bernal, G.; Duchaud, E.; Durand, L.; Dussurget, O.; Entian, K.D.; Feihl, H.; Science 294, 849-852, 2001  
 D.; Jones, L.M.; Karst, U.  
 A/Authors: Kretz, J.; Kunz, M.; Kunst, F.; Kurapkac, G.; Madueno, E.; Maitournam, A.; Maok, C.; Schluter, T.; Simoes, N.; Tixeret, A.; Vazquez-Boland, J.A.; Voss, H.; Wehland, A./Title: Comparative genomics of *Listeria species*.  
 A/Reference number: AB1077; MUID:21537279; PMID:11679669  
 A/Accession: AG1256  
 A/Status: preliminary  
 A/Molecule type: DNA

A/Residues: 1-626 <GUA>  
 A/Cross-references: GB:NC\_003210; PIDN:CA09533.1; PID:g16410884; GSPDB:GN00177  
 A/Experimental source: strain EGD-e  
 C/Genetics:  
 A/Gene: dnaG

Query Match 44.4%; Score 40; DB 2; Length 626;  
 Best Local Similarity 42.9%; Pred. No. 1.1e+02;  
 Matches 6; Conservative 4; Mismatches 4; Indels 0; Gaps 0;

OY 1 VGYGRDNDPRD 14  
 Db 527 IGFAEGNDADPTK 540

RESULT 33  
 T40107  
 hypothetical 57.9 kd protein - fission yeast (*Schizosaccharomyces pombe*)  
 C/Species: *Schizosaccharomyces pombe*  
 C/Date: 03-Dec-1999 #sequence\_revision 03-Dec-1999 #text\_change 03-Dec-1999  
 C/Accession: T40107  
 R/Wood, V.; Rajandream, M.A.; Barrell, B.G.; Taylor, K.; Harris, D. submitted to the EMBL Data Library, September 1998  
 A/Reference number: Z21906  
 A/Accession: T40107  
 A/Status: preliminary; translated from GB/EMBL/DDAT  
 A/Molecule type: DNA  
 A/Residues: 1-658 <MOO>  
 A/Cross-references: EMBL:AL031788; PIDN:CAA21162.1; GSPDB:GN00067; SPDB:SPBC2D10.04  
 A/Experimental source: strain 972n-; cosmid c2D10  
 C/Genetics:  
 A/Gene: SPDB:SPBC2D10.04  
 A/Map position: 2

Query Match 44.4%; Score 40; DB 2; Length 658;  
 Best Local Similarity 63.6%; Pred. No. 1.2e+02;  
 Matches 7; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

OY 2 GYGRDNDPRD 12  
 Db 26 GSSRDNDNSNP 36

RESULT 34  
 T24230  
 hypothetical protein R166.5 - *Caenorhabditis elegans*  
 C/Species: *Caenorhabditis elegans*  
 C/Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 15-Oct-1999  
 C/Accession: T24230  
 R/Mathews, P. submitted to the EMBL Data Library, August 1995  
 A/Reference number: Z19859  
 A/Accession: T24230  
 A/Status: preliminary; translated from GB/EMBL/DDAT  
 A/Molecule type: DNA  
 A/Residues: 1-761 <WIL>  
 A/Cross-references: EMBL:Z50795; PIDN:CAA90665.1; GSPDB:GN00020; CESP:R166.5  
 A/Experimental source: clone R166  
 C/Genetics:  
 A/Gene: CESP:R166.5  
 A/Map position: 2  
 A/Introns: 41/1; 198/1; 427/2; 535/2; 613/2; 747/3

Query Match 44.4%; Score 40; DB 2; Length 761;  
 Best Local Similarity 25.6%; Pred. No. 1.4e+02;  
 Matches 10; Conservative 1; Mismatches 4; Indels 24; Gaps 1;

OY 2 GYGRDNDPRD-----PSRKN 16  
 Db 152 GYGRDNDPRD-----PSRKN 190

RESULT 35

H83273  
 ribonuclease E PA2976 [imported] - Pseudomonas aeruginosa (strain PA01)  
 C/Species: Pseudomonas aeruginosa  
 C/Date: 15-Sep-2000 #sequence\_revision 15-Sep-2000 #text\_change 31-Dec-2000  
 C/Accession: H83273  
 R/Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warrenner, P.; Hickey, M.J.; Br  
 adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Lardig, K.; Lim,  
 .; Lory, S.; Olson, M.V.  
 Nature 406, 959-964, 2000  
 A/Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic patho  
 A/Reference number: A82950; MUID:20437337; PMID:10984043  
 A/Accession: H83273  
 A/Status: preliminary  
 A/Molecule type: DNA  
 A/Residues: 1-1057 <STO>  
 A/Cross-references: GB:AE004723; GB:AE004091; NID:g9949067; PIDN:AG06364.1; GSPDB:GN001  
 A/Experimental source: strain PA01  
 C/Genetics:  
 A/Gene: rne; PA2976  
 C/Superfamily: ribonuclease E

Query Match 44.4%; Score 40; DB 2; Length 1057;  
 Best Local Similarity 66.7%; Pred. No. 2e+02;  
 Matches 8; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

OY 4 GRDNDNDPDRSKN 15  
 ||| ||| |||  
 Db 604 GRDGNRDERK 615

RESULT 36  
 A45555  
 glutamate rich protein - malaria parasite (Plasmodium falciparum)  
 C/Species: Plasmodium falciparum  
 C/Date: 17-Feb-1994 #sequence\_revision 17-Feb-1994 #text\_change 09-Jun-2000  
 C/Accession: A45555; S27831  
 R/Borre, M.B.; Dziegiel, M.; Hogh, B.; Petersen, E.; Rieneck, K.; Riley, E.; Meis, J.F.;  
 J. Mol. Biochem. Parasitol. 49, 119-131, 1991  
 A/Title: Primary structure and localization of a conserved immunogenic Plasmodium falcip  
 rate life cycle.  
 A/Reference number: A45555; MUID:92131041; PMID:1775153  
 A/Accession: A45555  
 A/Status: preliminary  
 A/Molecule type: DNA  
 A/Residues: 1-1271 <BOR>  
 A/Cross-references: EMBL:M59706; NID:g160311; PID:g160312  
 A/Note: Sequence extracted from NCBI backbone (NCBI:77801, NCBIP:77802)

Query Match 44.4%; Score 40; DB 2; Length 1271;  
 Best Local Similarity 46.7%; Pred. No. 2.4e+02;  
 Matches 7; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

OY 2 GYGRDNDPDRSKN 16  
 ||| ||| |||  
 Db 1203 GLSRKDEKSSNNK 1217

RESULT 37  
 S25370  
 MSB2 protein - yeast (Saccharomyces cerevisiae)  
 C/Species: Saccharomyces cerevisiae  
 C/Date: 17-Apr-1993 #sequence\_revision 17-Apr-1993 #text\_change 21-Jul-2000  
 C/Accession: S25370; S64305  
 R/Bender, A.; Pringle, J.R.  
 Yeast 8, 315-323, 1992  
 A/Title: A Ser/Thr-rich multicopy suppressor of a cdc24 bud emergence defect.  
 A/Reference number: S25370; MUID:92383951; PMID:1514328  
 A/Accession: S25370  
 A/Status: preliminary  
 A/Molecule type: DNA  
 A/Residues: 1-1306 <BEN>  
 A/Cross-references: GB:M7354; NID:g171993; PIDN:AAA34798.1; PID:g171994

R/Rieger, M.; Mueller-Auer, S.; Brueckner, M.; Schaefer, M.  
 submitted to the Protein Sequence Database, May 1996  
 A/Reference number: S64071  
 A/Accession: S64305  
 A/Molecule type: DNA  
 A/Residues: 1-1306 <RIE>  
 A/Cross-references: EMBL:Z72799; NID:g1322977; PID:g1322978; MIPS:YGR014w  
 A/Experimental source: strain S288C  
 C/Genetics:  
 A/Gene: SGD:MSB2  
 A/Cross-references: SGD:S0003246; MIPS:YGR014w  
 A/Map position: 7R  
 C/Superfamily: yeast glucan 1,4-alpha-glucosidase homolog; glucan 1,4-alpha-glucosidase  
 C/Keywords: transmembrane protein  
 F/3-19/Domain: transmembrane #status predicted <TM1>  
 F/1189-1205/Domain: transmembrane #status predicted <TM2>

Query Match 44.4%; Score 40; DB 2; Length 1306;  
 Best Local Similarity 46.7%; Pred. No. 2.5e+02;  
 Matches 7; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

OY 2 GYGRDNDPDRSKN 16  
 ||| ||| |||  
 Db 1269 GHGLNNDSTPTRN 1283

RESULT 38  
 T25753  
 hypothetical protein F45E4.3 - Caenorhabditis elegans  
 C/Species: Caenorhabditis elegans  
 C/Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 15-Oct-1999  
 C/Accession: T25753  
 R/Milson, R.  
 submitted to the EMBL Data Library, September 1996  
 A/Description: The sequence of C. elegans cosmid F45E4.  
 A/Reference number: Z20082  
 A/Accession: T25753  
 A/Status: preliminary; translated from GB/EMBL/DBJ  
 A/Molecule type: DNA  
 A/Residues: 1-1325 <MIL>  
 A/Cross-references: EMBL:U70852; PIDN:AB09134.1; GSPDB:GN00022; CBSP:F45E4.3  
 A/Experimental source: strain Bristol N2; clone F45E4  
 C/Genetics:  
 A/Gene: CBSP:F45E4.3  
 A/Map position: 4  
 A/Intons: 25/3; 859/1; 928/1; 966/1; 1002/2; 1106/2; 1167/1; 1255/1; 1274/2

Query Match 44.4%; Score 40; DB 2; Length 1325;  
 Best Local Similarity 63.6%; Pred. No. 2.6e+02;  
 Matches 7; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

OY 6 DDNDPDRSKN 16  
 ||| ||| |||  
 Db 464 DDDDDPDRSKN 474

RESULT 39  
 G83424  
 hypothetical protein PA1764 [imported] - Pseudomonas aeruginosa (strain PA01)  
 C/Species: Pseudomonas aeruginosa  
 C/Date: 15-Sep-2000 #sequence\_revision 15-Sep-2000 #text\_change 31-Dec-2000  
 C/Accession: G83424  
 R/Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warrenner, P.; Hickey, M.J.; B  
 radman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Lardig, K.; Lim  
 .; Lory, S.; Olson, M.V.  
 Nature 406, 959-964, 2000  
 A/Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic path  
 A/Reference number: A82950; MUID:20437337; PMID:10984043  
 A/Accession: G83424  
 A/Status: preliminary  
 A/Molecule type: DNA  
 A/Residues: 1-532 <STO>  
 A/Cross-references: GB:AE004602; GB:AE004091; NID:g9947739; PIDN:AG05153.1; GSPDB:GN00

A;Experimental source: strain PA01  
C;Genetics:  
A;Gene: PA1764

Query Match	43.9%	Score 39.5	DB 2	Length 532
Best Local Similarity	57.1%	Pred. No. 1.2e+02		
Matches	8	Conservative	3	Mismatches 2
				Indels 1
				Gaps 1

```
QY      1 VGYGRDDNDRDPSR 14
      :|||:||:||
Db      157 LGYSR-DSDSDPAR 169
```

**RESULT 40**

**\$51129**

genome polypeptide - human rhinovirus 50 (fragment)

N;Contains: coat protein VP2

C;species: human rhinovirus 50  
C;Date: 19-Mar-1997 #sequence

```
C:\Accession: S51129
#sequence_revision 30-Jan-1998 #revl_change 26-Aug-1993
```

R;Horsnell, C.; Hug

submitted to the EMBL Data Library, January 1995

A; Description: Sequence analysis of 21 human rh

A;Reference number: S51121

A;Accession: S51129

A;Molecule type: genomic RNA

A;resIdues: 1-112 <HOK>  
A;Cross-references: EMBL:247

```
C:\Superfamily: poliovir
```

**C;Keywords:** coat protein; polyprotein

F;1-112/Product: coat protein VP2 (fr

1. **Introduction**

Query Match	Score	DB	Length
43.3%	39	2	112

Best Local Similarity 58.3%; Pred. No. 26;  
Matches 7; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

```
QY      5 RDDNDRDPSRKN 16
          |||:| |
Db      67 RDQNDKQPSDDN 78
```

Search completed: April 9, 2003, 12:59:01  
Job time : 9.95082 secs

GenCore version 5.1.4 p5 4578  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using SW model

Run on: April 9, 2003, 12:40:41 ; Search time 3.80328 Seconds  
(without alignments)  
174.487 Million cell updates/sec

Title: US-10-020-441-3

Perfect score: 90  
Sequence: 1 VGYGRDDNDPDRSKN 16

Scoring table: BLOSUM62  
Gapop 10.0, Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%  
Listing first 45 summaries

Database: SwissProt\_40.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	90	100.0	264	1 CERC_SCMA	P12546 schistosoma
2	48.5	53.9	462	1 YMO1_YEAST	Q04461 saccharomyc
3	46	51.1	1428	1 TOP2_YEAST	P06786 saccharomyc
4	44	48.9	420	1 METK_AMOPR	O09486 amoeba prot
5	42	46.7	655	1 YKDA_MYCCA	P45615 mycoplasma
6	42	46.7	1382	1 IF3A_HUMAN	Q14162 homo sapien
7	42	46.7	2559	1 FAFK_MOUSE	P70399 m probabie
8	41	45.6	427	1 RUI7_ARATH	Q42404 arabidopsis
9	41	45.6	789	1 CAD9_HUMAN	Q01034 homo sapien
10	41	45.6	1406	1 TOP1_CANGA	O93794 candida gla
11	40.5	45.0	829	1 IF2_HAETN	P44323 haemophilus
12	40	44.4	192	1 Y045_NPVAC	P41450 autographa
13	40	44.4	355	1 SDC_DROME	P49415 drosophila
14	40	44.4	496	1 U2AF_CABEL	P09978 caenorhabdi
15	40	44.4	626	1 PRIM_LISMO	P47762 listeria mo
16	40	44.4	1306	1 MSB2_YEAST	P23334 saccharomyc
17	39.5	43.9	712	1 ARS2_CABEL	Q06615 caenorhabdi
18	39	43.3	146	1 CYM5_RAT	P04166 ratus norv
19	39	43.3	162	1 CAV2_HUMAN	P51666 homo sapien
20	39	43.3	206	1 YEEN_ECOLI	P76264 escherichia
21	39	43.3	534	1 NAB4_YEAST	Q39383 saccharomyc
22	39	43.3	555	1 HEXA_CABEL	Q22492 caenorhabdi
23	39	43.3	615	1 YSPK_CABEL	Q19425 caenorhabdi
24	39	43.3	739	1 DD15_CABEL	Q10875 caenorhabdi
25	39	43.3	1270	1 VGLN_CHICK	P01021 gallus gall
26	39	43.3	1319	1 SSM4_YEAST	P00318 saccharomyc
27	38.5	42.8	397	1 CPXQ_SACER	P48635 saccharopol
28	38	42.2	244	1 YL53_CABEL	P34433 caenorhabdi
29	38	42.2	305	1 PIX1_XENTLA	Q07751 xenopus lae
30	38	42.2	311	1 CAT4_ACTICA	P07743 actinobact
31	38	42.2	337	1 DJB4_HUMAN	Q04833 homo sapien
32	38	42.2	337	1 DJB4_MOUSE	Q04833 mus musculu
33	38	42.2	462	1 SEM1_HUMAN	P04279 homo sapien

34	38	42.2	514	1 WRK4_ARATH	Q9x190 arabidopsis
35	38	42.2	530	1 AHPF_XANCH	O06465 xanthomonas
36	38	42.2	560	1 TATR_NPVOP	P22114 oryza pseu
37	38	42.2	564	1 FMRI_XENTLA	P51113 xenopus lae
38	38	42.2	594	1 DBP1_YEAST	O06218 saccharomyc
39	38	42.2	626	1 PRIM_LISIN	Q92505 listeria in
40	38	42.2	659	1 YHHS_YEAST	P38760 saccharomyc
41	38	42.2	755	1 COMP_RAT	P35444 ratus norv
42	38	42.2	1253	1 DSPP_HUMAN	Q9n244 homo sapien
43	38	42.2	1310	1 YB35_SCHPO	O14340 schizosach
44	38	42.2	1446	1 IE18_PRYKA	P33479 pseudorabie
45	38	42.2	1461	1 IE18_PRYTF	P11675 pseudorabie

## ALIGNMENTS

RESULT 1	CERC_SCMA	STANDARD;	PRT;	264 AA.
AC	P12546;			
DT	01-OCT-1989 (Rel. 12, Created)			
DT	01-OCT-1989 (Rel. 12, Last sequence update)			
DT	15-JUN-2002 (Rel. 41, Last annotation update)			
DE	Cercarial protease precursor (EC 3.4.21.-) (Cercarial elastase).			
OS	Schistosoma mansoni (Blood fluke).			
OC	Eukaryota; Metazoa; Platyhelminthes; Trematoda; Digenea; Strigeidida;			
OC	Schistosomatidae; Schistosomatidae; Schistosoma.			
OX	NCBI_Taxid=6183;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE=8830818; PubMed=3166457;			
RA	Newport G.R., McKerrrow J.H., Hedstrom R., Pettit M., McGarrigle L.,			
RA	Bart P.J., Agabian N.			
RT	"Cloning of the proteinase that facilitates infection by schistosome			
RT	parasites."			
RT	J. Biol. Chem. 263.13179-13184(1988).			
CC	-1- FUNCTION: THIS PROTEASE CLEAVES ELASTIN AND THUS FACILITATES			
CC	PENETRATION OF SCHISTOSOME PARASITE LARVAE THROUGH ELASTIN-			
CC	RICH TISSUE OF THE HOST.			
CC	-1- ENZYME REGULATION: ACTIVATED BY AN AUTOCATALYTIC MECHANISM.			
CC	-1- SUBCELLULAR LOCATION: ACETABULAR (PENETRATION) GLANDS.			
CC	-1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1.			
CC	-----			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration			
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -			
CC	the European Bioinformatics Institute. There are no restrictions on its			
CC	use by non-profit institutions as long as its content is in no way			
CC	modified and this statement is not removed. Usage by and for commercial			
CC	entities requires a license agreement (See <a href="http://www.isb-sib.ch/announce/">http://www.isb-sib.ch/announce/</a>			
CC	or send an email to <a href="mailto:license@isb-sib.ch">license@isb-sib.ch</a> ).			
CC	-----			
DR	EMBL; J03946; AAA29864.1; -			
DR	PIR; A28942; A28942.			
DR	HSSP; P20231; IAAO.			
DR	MEROB; S01.144; -			
DR	InterPro; IPR001314; Chymotrypsin.			
DR	InterPro; IPR001254; Ser_Protease_Try.			
DR	Pfam; PF00089; Trypsin_1.			
DR	PRINTS; PR00722; CHYMOTRYPSIN.			
DR	SMART; SM00020; TRY_PSP; 1.			
DR	PROSITE; PS50240; TRYPSIN_DOM; 1.			
DR	PROSITE; PS00134; TRYPSIN_HTS; 1.			
DR	PROSITE; PS00135; TRYPSIN_SER; FALSE_NEG.			
DR	KW Hydroxylase; Serine protease; Zymogen; Signal.			
FT	SIGNAL	1	19	POTENTIAL.
FT	PROPEP	20	27	POTENTIAL.
FT	CHAIN	28	264	CERCARIAL PROTEASE.
FT	ACT_SITE	68	68	CHARGE RELAY SYSTEM (BY SIMILARITY).
FT	ACT_SITE	126	126	CHARGE RELAY SYSTEM (BY SIMILARITY).
FT	ACT_SITE	218	218	CHARGE RELAY SYSTEM (BY SIMILARITY).
SQ	SEQUENCE	264 AA;	28545 MW;	E2E5129A7C5D5010 CRC64;



```

FT MOD_RES 1272 1272 PHOSPHORYLATION (BY CK2).
FT MOD_RES 1353 1353 PHOSPHORYLATION (BY CK2).
FT MOD_RES 1356 1356 PHOSPHORYLATION (BY CK2).
FT MOD_RES 1408 1408 PHOSPHORYLATION (BY CK2).
FT MOD_RES 1423 1423 PHOSPHORYLATION (BY CK2).
FT CONFLICT 74 74 N -> NN (IN REF. 1).
FT CONFLICT 547 547 P -> L (IN REF. 1).
FT CONFLICT 837 837 W -> R (IN REF. 1).
SQ SEQUENCE 1426 AA; 164214 MW; 0E29EB8B93A1387 CRC64;

Query Match
Best Local Similarity 51.1%; Score 46; DB 1; Length 1428;
Matches 8; Conservative 1; Mismatches 2; Indels 0; Gaps 0;
QY 6 DDNDRDPSSRN 16
DB 1198 DDEDYDPSKXN 1208

```

## RESULT 4

```

METK_AMOPR STANDARD; PRT; 420 AA.
AC 009486;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE S-adenosylmethionine synthetase (EC 2.5.1.6) (Methionine
adenosyltransferase) (Adomet synthetase) (Fragment).
OS Amoeba proteus (Amoeba).
OC Eukaryota; Lobosea; Eumeboida; Amoebidae; Amoeba.
OX NCBI_TaxId=5775;
RN [1]
RP SEQUENCE FROM N.A.
RC MEDLINE=97449794; PubMed=9304810;
RA Choi J.Y., Lee T.W., Jeon K.W., Ahn T.I.;
RT "Evidence for symbiont-induced alteration of a host's gene expression:
irreversible loss of SAM synthetase from Amoeba proteus.";
RL J. Eukaryot. Microbiol. 44:412-419(1997).
CC -1- FUNCTION: CATALYZES THE FORMATION OF S-ADENOSYLMETHIONINE FROM
METHIONINE AND ATP (BY SIMILARITY).
CC -1- CATALYTIC ACTIVITY: ATP + L-methionine +
diphosphate + S-adenosyl-L-methionine.
CC -1- PATHWAY: Activated methyl cycle.
CC -1- SUBCELLULAR LOCATION: Cytoplasmic.
CC -1- SIMILARITY: BELONGS TO THE ADOMET SYNTHETASE FAMILY.
CC -----
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
CC -----
CC EMBL: U91602; AAB50563.1; -.
DR HSSP; P04384; IMXB.
DR InterPro: IPR002133; S-Adomet_synth.
DR Pfam; PF00438; S-Adomet_synth. 1.
DR Pfam; PF02772; S-Adomet_synth2; 1.
DR Pfam; PF02773; S-Adomet_synth3; 1.
DR TIGRPFAM; TIGR01034; metK; 1.
DR PROSITE; PS00376; ADOMET_SYNTHETASE_1; 1.
DR PROSITE; PS00377; ADOMET_SYNTHETASE_2; 1.
KW transferase; One-carbon metabolism; ATP-binding.
FT NON TER 1 1
FT NP BIND 120 125 ATP (POTENTIAL).
SQ SEQUENCE 420 AA; 47161 MW; 1CAF0CFDD26345C7 CRC64;

```

```

Query Match
Best Local Similarity 48.9%; Score 44; DB 1; Length 420;
Matches 7; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

```

```

QY 1 VGYGRDNDSDPS 13
DB 77 IGYSRDLDIDPRT 89

```

## RESULT 5

```

YKDA_MYCCA STANDARD; PRT; 655 AA.
AC P45615;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical 75.9 kDa protein in Kd1B 5' region (ORFA).
OS Mycoplasma capricolum.
OC Bacteria; Firmicutes; Mollicutes; Entomoplasmatales;
OC Entomoplasmataceae.
OX NCBI_TaxId=2095;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 27343 / K1D;
RX MEDLINE=95218642; PubMed=7703858;
RA Zhu P.-P., Reizer J., Peterkofsky A.;
RT "Unique distal operon (ptsI-cir) in Mycoplasma capricolum
encoding enzyme I and the glucose-specific enzyme IIA of the
phosphoenolpyruvate:sugar phosphotransferase system: cloning,
RT sequencing, promoter analysis, and protein characterization.";
RL Protein Sci. 3:2115-2128(1994).
CC -1- SIMILARITY: BELONGS TO THE MG067 / MG068 / MG395 FAMILY.
CC -----
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
CC -----
CC EMBL: U15110; AAA70404.1; -.
DR InterPro: IPR002414; DUF30/31.
DR Pfam; PF01732; DUF31; 1.
DR PRINTS; PR00840; Y06768FAMILY.
KW Hypothetical protein.
SQ SEQUENCE 655 AA; 75908 MW; AD3E4B0EAD07B060 CRC64;

```

```

Query Match
Best Local Similarity 46.7%; Score 42; DB 1; Length 655;
Matches 9; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

```

```

QY 1 VGYGRDNDSDPSRN 16
DB 277 LGLGKADNVDFSRKN 292

```

## RESULT 6

```

IF3A_HUMAN STANDARD; PRT; 1382 AA.
AC Q14152; O00653;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE Eukaryotic translation initiation factor 3 subunit 10 (eIF-3 theta)
DE (EIF3 p167) (eIF3 p180) (eIF3 p185).
GN EIF3S10 OR KIAA0139.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxId=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Bone marrow.
RX MEDLINE=96177530; PubMed=8590280;
RA Nagase T., Seki N., Tanaka A., Ishikawa K.-I., Nomura N.;
RT "Prediction of the coding sequences of unidentified human genes. IV.

```

RT The coding sequences of 40 new genes (K1AA0121-K1AA0160) deduced by  
 RT analysis of cDNA clones from human cell line KG-1.";  
 RL DNA Res. 2:167-174(1995).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=97294683; PubMed=9150439;  
 RA Scholler J.K., Kanner S.B.;  
 RT "The human p167 gene encodes a unique structural protein that contains  
 RT centrosomal A homology and associates with a multicomponent complex.";  
 RL DNA Cell Biol. 16:515-531(1997).  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=97207269; PubMed=9054404;  
 RA Johnson K.R., Merrick W.C., Zoll W.L., Zhu Y.;  
 RT "Identification of cDNA clones for the large subunit of eukaryotic  
 RT translation initiation factor 3. Comparison of homologues from human,  
 RT Nicotiana tabacum, Caenorhabditis elegans, and Saccharomyces  
 RT cerevisiae.";  
 RL J. Biol. Chem. 272:7106-7113(1997).  
 CC -1- FUNCTION: BINDS TO THE 40S RIBOSOME AND PROMOTES THE BINDING OF  
 CC METHIONYL-TRNAI AND MRNA.  
 CC -1- SUBCELLULAR LOCATION: Cytoplasmic.  
 CC -1- PPM: PHOSPHORYLATED.  
 CC -1- SIMILARITY: BELONGS TO THE EIF3S10 FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; D50929; BAA09488.1; -;  
 DR EMBL; U58046; AAB41584.1; -;  
 DR EMBL; U78311; AAB80695.1; -;  
 DR Genew; HGNC:3271; EIF3S10.  
 DR MIM; 602039; -;  
 DR InterPro; IPR002017; PCI.  
 DR InterPro; IPR002017; Spectrin.  
 DR Pfam; PF01399; PCI.1.  
 DR SMART; SM00088; PINT.1.  
 KM Initiation factor; Protein biosynthesis; Repeat; Phosphorylation.  
 FT DOMAIN 925 1172  
 FT 25 X 10 AA TANDEM REPEAT OF D-[DE]-D-R-  
 FT [GP]-[PS]-[RW]-R-[GN]-[AM].  
 SQ SEQUENCE 1382 AA; 166568 MW; 485C01B28067EBBA CRC64;  
 Query Match 46.7%; Score 42; DB 1; Length 1382;  
 Best Local Similarity 54.5%; Pred. No. 61;  
 Matches 6; Conservative 4; Mismatches 1; Indels 0; Gaps 0;  
 QY 5 RDDNDPDSRK 15  
 DB 1216 REENDKDPERE 1226  
 RESULT 7  
 FAFX\_MOUSE STANDARD; PRT; 2559 AA.  
 AC P70398; O62497;  
 DT 01-NOV-1997 (Rel. 35, Created)  
 DT 01-NOV-1997 (Rel. 35, Last sequence update)  
 DT 30-MAY-2000 (Rel. 39, Last annotation update)  
 DE Probable ubiquitin carboxyl-terminal hydrolase FAF-X (EC 3.1.2.15)  
 DE (Ubiquitin thiolesterase FAF-X) (Ubiquitin-specific processing  
 DE protease FAF-X) (Deubiquitinating enzyme FAF-X) (Fat facets protein  
 DE related, X-linked) (Ubiquitin-specific protease 9, X chromosome)  
 DE (Ubiquitin carboxyl-terminal hydrolase FAM) (Fat facets homolog).  
 GN USP9X OR FAM OR FAFU.  
 OS Mus musculus (Mouse).  
 CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 CC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.

OX NCBI\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=97321544; PubMed=9178254;  
 RA Wood S.A., Pascoe W.S., Ru K., Yamada T., Hirschman J., Kemler R.,  
 RA Matlack J.S.;  
 RT "Cloning and expression analysis of a novel mouse gene with sequence  
 RT similarity to the Drosophila fat facets gene.";  
 RL Mech. Dev. 63:29-38(1997).  
 RN [2]  
 RP SEQUENCE OF 152-240 FROM N.A.  
 RC TISSUE=Cochlea;  
 RX MEDLINE=97237053; PubMed=9119401;  
 RA Crozet F., Amraoui A.E., Blanchard S., Lenoir M., Ripoll C., Vago P.,  
 RA Hamel C., Fzames C., Levi-Acobas F., Depetris D., Mattei M.-G.,  
 RA Weil D., Pujol R., Petit C.;  
 RT "Cloning of the genes encoding two murine and human cochlear  
 RT unconventional type I myosins.";  
 RL Genomics 40:332-341(1997).  
 CC -1- FUNCTION: MAY FUNCTION AS AN UBIQUITIN-PROTEIN OR POLYUBIQUITIN  
 CC AND OF UBIQUITINATED PROTEINS.  
 CC HYDROLASE INVOLVED BOTH IN THE PROCESSING OF UBIQUITIN PRECURSORS  
 CC AND OF UBIQUITINATED PROTEINS.  
 CC -1- CATALYTIC ACTIVITY: Ubiquitin C-terminal thiolester + H(2)O =  
 CC ubiquitin + a thiol.  
 CC -1- TISSUE SPECIFICITY: UBIQUITOUSLY EXPRESSED IN ADULT TISSUES.  
 CC -1- DEVELOPMENTAL STAGE: AT LEAST EXPRESSED FROM 17 DAYS POST-CONITUM  
 CC TO 21 POSTNATAL DAYS.  
 CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY C19.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; U67874; AAB07731.1; -;  
 DR EMBL; Z78153; CAB01555.1; -;  
 DR MEROPS; C19.017; -;  
 DR MGD; MGI:894681; Usp9x.  
 DR InterPro; IPR001394; UCH-2.  
 DR Pfam; PF00442; UCH-1; 1.  
 DR Pfam; PF00443; UCH-2; 1.  
 DR PROSITE; PS00972; UCH\_2\_1; 1.  
 DR PROSITE; PS00973; UCH\_2\_2; 1.  
 DR PROSITE; PS50235; UCH\_2\_3; 1.  
 KM Ub1 conjugation pathway; Hydrolase; Thiol protease; Multigene family.  
 FT ACT\_SITE 1566 1566  
 FT ACT\_SITE 1871 1871  
 FT ACT\_SITE 1879 1879  
 FT ACT\_SITE 182 182  
 FT CONFLICT 182 182  
 FT C -> F (IN REF. 2).  
 SQ SEQUENCE 2559 AA; 290542 MW; 9BDF964AA142AFB5D CRC64;  
 Query Match 46.7%; Score 42; DB 1; Length 2559;  
 Best Local Similarity 70.0%; Pred. No. 1.2e+02;  
 Matches 7; Conservative 1; Mismatches 2; Indels 0; Gaps 0;  
 QY 3 YGRDNDPDP 12  
 DB 630 YGRDNDYDP 639  
 RESULT 8  
 RUT7\_ARATH STANDARD; PRT; 427 AA.  
 AC Q42404; Q42378;  
 DT 16-OCT-2001 (Rel. 40, Created)  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE U1 small nuclear ribonucleoprotein 70 kDa (U1 snRNP 70 kDa) (snRNP70)  
 DE (U1-70K).  
 GN RNU1 OR AT3G50670 OR T3A5.50.





KW Cell adhesion; Glycoprotein; Transmembrane; Calcium-binding; Repeat;  
 FT SIGNAL 1 21 POTENTIAL.  
 FT PROPEP 22 53 POTENTIAL.  
 FT CHAIN 54 789 CADHERIN-9.  
 FT DOMAIN 54 615 EXTRACELLULAR (POTENTIAL).  
 FT TRANSMEM 616 636 POTENTIAL.  
 FT DOMAIN 637 789 CYTOPLASMIC (POTENTIAL).  
 FT DOMAIN 54 159 CADHERIN 1.  
 FT DOMAIN 160 268 CADHERIN 2.  
 FT DOMAIN 269 383 CADHERIN 3.  
 FT DOMAIN 384 486 CADHERIN 4.  
 FT DOMAIN 487 608 CADHERIN 5.  
 FT CARBOHYD 255 255 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 437 437 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 455 455 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 536 536 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 SQ SEQUENCE 789 AA; 88701 MW; 8598532756A4344F CRC64;

Query Match 45.6%; Score 41; DB 1; Length 789;  
 Best Local Similarity 77.8%; Pred. No. 48;  
 Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 3 YGRDNDNRD 11  
 |||:|||||  
 Db 781 YGDDSDRD 789

## RESULT 10

TOP1\_CANGA STANDARD; PRT; 1406 AA.

AC 093794;

DT 15-JUN-2002 (Rel. 41, Created)

DT 15-JUN-2002 (Rel. 41, Last sequence update)

DT 15-JUN-2002 (Rel. 41, Last annotation update)

DE DNA topoisomerase II (EC 5.99.1.3).

GN TOP2.

OS Candida glabrata (Yeast) (Torulopsis glabrata).

OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;

OC Saccharomycetales; mitosporic Saccharomycetales; Candida.

OX NCBI\_TaxId=5478;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=ATCC 2001;

RX MEDLINE=98455808; PubMed=9782488;

RA Nayayama H., Izura M., Nagahashi S., Sinha E.Y., Sato Y., Yamazaki T.,

RA Aisawa M., Kitada K.,

RT "A controllable gene-expression system for the pathogenic fungus

Candida glabrata.";

RL Microbiology 144:2407-2415(1998).

CC -1- FUNCTION: CONTROL OF TOPOLOGICAL STATES OF DNA BY TRANSIENT

BREAKAGE AND SUBSEQUENT REJOINING OF DNA STRANDS. TOPOISOMERASE II

MAKES DOUBLE-STRAND BREAKS (By similarity).

-1- CATALYTIC ACTIVITY: ATP-dependent breakage, passage and rejoining

of double-stranded DNA.

-1- SUBUNIT: HOMODIMER (By similarity).

-1- SUBCELLULAR LOCATION: Nuclear (By similarity).

-1- MISCELLANEOUS: EUKARYOTIC TOPOISOMERASE I AND II CAN RELAX BOTH

NEGATIVE AND POSITIVE SUPERCOILS, WHEREAS PROKARYOTIC ENZYMES

RELAX ONLY NEGATIVE SUPERCOILS.

-1- SIMILARITY: BELONGS TO THE TYPE II TOPOISOMERASE FAMILY.

-----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

between the Swiss Institute of Bioinformatics and the EMBL outstation -

the European Bioinformatics Institute. There are no restrictions on its

use by non-profit institutions as long as its content is in no way

modified and this statement is not removed. Usage by and for commercial

entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

DR InterPro; IPR003957; CBFA\_NFYB\_topis.  
 DR InterPro; IPR001241; DNA\_topoisol1.  
 DR InterPro; IPR002205; DNA\_topoisol1.  
 DR Pfam; PF00204; DNA\_gyraseB.1.  
 DR Pfam; PF00521; DNA\_topoisol1V.1.  
 DR Pfam; PF02518; HATPase\_C.1.  
 DR PRINTS; PR00615; CCAATSUBONTA.  
 DR PRINTS; PR00418; TP12FAMILY.  
 DR PRODOM; PD000742; DNA\_topoisol1V.1.  
 DR SMART; SM00387; HATPase\_C.1.  
 DR SMART; SM00433; TOP2C.1.  
 DR SMART; SM00434; TOP4C.1.  
 DR PROSITE; PS00177; TOPOISOMERASE\_II.1.  
 KW Isomerase; Topoisomerase; DNA-binding; ATP-binding; Phosphorylation;  
 KW Nuclear protein.  
 FT NP BIND 139 144 ATP (POTENTIAL).  
 FT ACT SITE 780 780 DNA CLEAVAGE (By similarity).  
 SQ SEQUENCE 1406 AA; 161017 MW; 97A1CD6B49A5DD91 CRC64;

Query Match 45.6%; Score 41; DB 1; Length 1406;  
 Best Local Similarity 66.7%; Pred. No. 91;  
 Matches 8; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

Qy 4 GRDNDPRDPRK 15  
 |||:|||||  
 Db 1193 GADDDYDPSGK 1204

## RESULT 11

IF2\_HAEIN STANDARD; PRT; 829 AA.

AC P44323;

DT 01-NOV-1995 (Rel. 32, Created)

DT 01-NOV-1995 (Rel. 32, Last sequence update)

DT 15-JUN-2002 (Rel. 41, Last annotation update)

DE Translation initiation factor IF-2.

GN INF OR H11284.

OS Haemophilus influenzae.

OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;

OC Haemophilus.

OX NCBI\_TaxId=727;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=Rd / KM20 / ATCC 51907;

RX MEDLINE=95350630; PubMed=7542800;

RA Fleischmann R.D., Adams M.D., White O., Clayton R.A., Kirkness E.F.,

RA Kerlavage A.R., Bult C.J., Tomb J.-F., Dougherty B.A., Merrick J.M.,

RA McKenney K., Sutton G., Fitzhugh W., Fields C.A., Gocayne J.D.,

RA Scott J.D., Shirley R., Liu L.-I., Glodek A., Kelley J.M.,

RA Weidman J.F., Phillips C.A., Spriggs T., Hedblom E., Cotton M.D.,

RA Uettermann T.R., Hanna M.C., Nguyen D.T., Saudek D.M., Brandon R.C.,

RA Gish W.L., Fritchman J.L., Fuhmann J.L., Geoghegan N.S.M.,

RA Venter J.C., McDonald L.A., Small K.V., Fraser C.M., Smith H.O.,

RT Venter J.C.;

RL "Whole-genome random sequencing and assembly of Haemophilus influenzae

Rd.";

CC Science 269:496-512(1995).

CC -1- FUNCTION: IF-2, ONE OF THE ESSENTIAL COMPONENTS FOR THE INITIATION

OF PROTEIN SYNTHESIS IN VITRO, PROTECTS FORMYL METHIONYL-TRNA FROM

SPONTANEOUS HYDROLYSIS AND PROMOTES ITS BINDING TO THE 30S

RIBOSOMAL SUBUNITS. IT IS ALSO INVOLVED IN THE HYDROLYSIS OF GTP

DURING THE FORMATION OF THE 70S RIBOSOMAL COMPLEX (By similarity).

-1- SIMILARITY: BELONGS TO THE IF-2 FAMILY.

-----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

between the Swiss Institute of Bioinformatics and the EMBL outstation -

the European Bioinformatics Institute. There are no restrictions on its

use by non-profit institutions as long as its content is in no way

modified and this statement is not removed. Usage by and for commercial

entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

```

DR EMBL; U32808; AAC22933.1; -.
DR TIGR; H11284; -.
DR InterPro; IPR004161; EFTU D2.
DR InterPro; IPR000795; EF_GTPbind.
DR InterPro; IPR000178; IF2.
DR InterPro; IPR005225; Small_GTP.
DR Pfam; PF00009; GTP_EFTU; 1.
DR Pfam; PF03144; GTP_EFTU D2; 2.
DR ProDom; PD186100; IF2; 1.
DR TIGRFAMs; TIGR00231; small_GTP; 1.
DR TIGRFAMs; TIGR00487; IF-2; 1.
DR PROSITE; PS01176; IF2; 1.
DR Initiation factor; Protein biosynthesis; GTP-binding;
KM Complete proteome.
FT DOMAIN 331 G-DOMAIN.
FT NP_BIND 337 344 GTP (BY SIMILARITY).
FT NP_BIND 384 388 GTP (BY SIMILARITY).
FT NP_BIND 441 GTP (BY SIMILARITY).
SQ SEQUENCE 829 AA; 90551 MW; 75B20AC4CF610AF7 CRC64;

Query Match 45.0%; Score 40.5; DB 1; Length 829;
Best Local Similarity 45.0%; Pred. No. 61;
Matches 9; Conservative 3; Mismatches 1; Indels 7; Gaps 1;

QY 4 GRDD-----NDRDPSRKN 16
Db 197 GRDDENSKNSKNSRKN 216

RESULT 12
Y045_NPVAC STANDARD; PRT; 192 AA.
AC P41450;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 01-NOV-1995 (Rel. 32, Last annotation update)
DE Hypothetical 22.7 kDa protein in GTP-P79 intergenic region.
OS Autographa californica nuclear polyhedrosis virus (ACNPV).
OC Viruses; dsDNA viruses, no RNA stage; Baculoviridae;
OC Nucleopolyhedrovirus.
OC NCBI_TaxID=46015;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C6;
RA MEDLINE=94303173; PubMed=8030224;
RT "Ayres M.D., Howard S.C., Kuzio J., Lopez-Ferber M., Possee R.D.;
RT "The complete DNA sequence of Autographa californica nuclear
RT polyhedrosis virus.";
RL Virology 202;586-605(1994).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; L22858; AAA66675.1; -.
DR Hypothetical protein.
SQ SEQUENCE 192 AA; 22651 MW; B2CDD9ACC247E2A7 CRC64;

Query Match 44.4%; Score 40; DB 1; Length 192;
Best Local Similarity 87.5%; Pred. No. 15;
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 5 RDDNDRDP 12
Db 50 RDDNDDB 57

RESULT 13
SDC_DROME

```

```

ID SDC_DROME STANDARD; PRT; 395 AA.
AC P49415;
DT 01-FEB-1996 (Rel. 33, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE Syndecan precursor.
GN SDC OR SYD.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Mandibulata; Pancrustacea; Hexapoda;
OC Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
OC Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
OC NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=94211858; PubMed=8159748;
RA Spring J., Paine-Saunders S.E., Hynes R.O., Bernfield M.;
RT "Drosophila syndecan: conservation of a cell-surface heparan sulfate
RT proteoglycan.";
RL Proc. Natl. Acad. Sci. U.S.A. 91;3334-3338(1994).
CC -1- FUNCTION: CELL SURFACE PROTEOGLYCAN THAT BEAR HEPARAN SULFATE.
CC -1- SUBCELLULAR LOCATION: Type I membrane protein.
CC -1- SIMILARITY: BELONGS TO THE SYNDECAN FAMILY OF INTEGRAL MEMBRANE
CC PROTEOGLYCANS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; U03282; AAC34307.1; -.
DR FlyBase; FBgn0010415; Sdc.
DR InterPro; IPR003585; Neurexin-like.
DR InterPro; IPR01050; Syndecan.
DR Pfam; PF01034; Syndecan; 1.
DR SMART; SM00294; 4.tn; 1.
DR PROSITE; PS00964; SYNDECAN; 1.
KM Proteoglycan; Heparan sulfate; Transmembrane; Glycoprotein; Signal.
FT SIGNAL 1 28
FT CHAIN 29 395
FT DOMAIN 29 335
FT TRANSMEM 336 360
FT DOMAIN 361 395
FT DOMAIN 114 169
FT CARBOHYD 62 62
FT CARBOHYD 79 79
FT CARBOHYD 81 81
FT CARBOHYD 110 110
FT CARBOHYD 155 155
FT CARBOHYD 190 190
SQ SEQUENCE 395 AA; 41853 MW; 4FF4AE3952240162 CRC64;

Query Match 44.4%; Score 40; DB 1; Length 395;
Best Local Similarity 46.2%; Pred. No. 33;
Matches 6; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

QY 2 GYGRDNDNDPSR 14
Db 82 GYGRDNDDBDPQ 94

RESULT 14
UZAF_CAEEL STANDARD; PRT; 496 AA.
ID UZAF_CAEEL
AC P90978; O76839; Q95XTO; Q9N2X5;
DT 15-DEC-1998 (Rel. 37, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Splicing factor UZAF 65 kDa subunit (U2 auxiliary factor 65 kDa
DE subunit) (U2 snRNP auxiliary factor large subunit) (UZAF65).
GN UAF-1 OR Y92C3B.2.

```

OS Caenorhabditis elegans.  
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;  
 OC Rhabditidae; Peioderinae; Caenorhabditis.  
 OX NCBI\_TaxID=6239;  
 RN [1]  
 RP SEQUENCE FROM N.A. (ISOFORM A).  
 RC STRAIN=Bristol N2; TISSUE=Embryo;  
 RX MEDLINE=97154530; PubMed=9001248;  
 RA Zorio D.A.R., Lea K., Blumenthal T.;  
 RT "Cloning of *Caenorhabditis* U2AF65: an alternatively spliced RNA  
 containing a novel exon.";  
 RL Mol. Cell. Biol. 17:946-953 (1997).  
 RN [2]  
 RP REVISIONS.  
 RA Zorio D.A.R.;  
 RL Submitted (JUL-1998) to the EMBL/GenBank/DBJ databases.  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Bristol N2;  
 RA Lamar B.;  
 RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.  
 RN [4]  
 RP REVISIONS. AND ALTERNATIVE SPLICING.  
 RA Waeberon R.;  
 RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.  
 CC -1- FUNCTION: NECESSARY FOR THE SPLICING OF PRE-MRNA. BINDS TO THE  
 POLYPYRIMIDINE TRACT OF INTRONS EARLY DURING SPLICOSOME ASSEMBLY  
 (BY SIMILARITY).  
 CC -1- SUBUNIT: FORMS A HETERODIMER WITH THE U2AF SMALL SUBUNIT (BY  
 SIMILARITY).  
 CC -1- SUBCELLULAR LOCATION: Nuclear.  
 CC -1- ALTERNATIVE PRODUCTS: 3 isoforms; a (shown here), b and c; may be  
 produced by alternative splicing.  
 CC -1- SIMILARITY: CONTAINS 3 RNA RECOGNITION MOTIFS (RRM).  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
 or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; U79157; AAC26982.1; -;  
 DR EMBL; AC024875; AAK29989.1; -;  
 DR EMBL; AC024875; AAK29990.1; -;  
 DR EMBL; AC024875; AAL00879.1; -;  
 DR WormPep; Y92C3B.2a; CE27339.  
 DR WormPep; Y92C3B.2b; CE25626.  
 DR WormPep; Y92C3B.2c; CE29149.  
 DR HSP; P26368; IJZF.  
 DR InterPro: IPR000504; RNA\_rec\_mot.  
 DR Pfam; PF00076; rrm; 3.  
 DR SMART; SM00360; RRM; 3.  
 DR PROSITE; PS50102; RRM; 3.  
 DR PROSITE; PS00030; RRM\_RNP\_1; 1.  
 KW Nuclear protein; RNA-binding; mRNA splicing; Repeat;  
 KW Alternative splicing.  
 FT DOMAIN 57 102 ARG/SER-RICH (RS DOMAIN).  
 FT DOMAIN 184 266 RNA-BINDING (RRM) 1.  
 FT DOMAIN 291 368 RNA-BINDING (RRM) 2.  
 FT DOMAIN 404 488 RNA-BINDING (RRM) 3.  
 FT DOMAIN 48 52 POLY-GLY.  
 FT DOMAIN 85 90 POLY-GLY.  
 FT DOMAIN 103 106 POLY-GLY.  
 FT VARSPLIC 1 353 MISSING (IN ISOFORM B).  
 FT VARSPLIC 57 78 MISSING (IN ISOFORM C).  
 SQ SEQUENCE 496 AA; 55430 MM; 1A967DE89DC9373C CRC64;

QY 2 GYGRDNDPSPSRKN 16  
 Db 50 GCGEDNDPDKRRKRS 64  
 RESULT 15  
 ID PRIM\_LISMO STANDARD; PRT; 626 AA.  
 AC P4762;  
 DT 01-FEB-1996 (Rel. 33, Created)  
 DT 01-FEB-1996 (Rel. 33, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE DNA primase (EC 2.7.7.-).  
 GN DNAG OR LMO1455.  
 OS Listeria monocytogenes.  
 OC Bacteria; Firmicutes; Bacillales; Listeriaceae; Listeria.  
 OX NCBI\_TaxID=1639;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=ATCC 13932;  
 RX MEDLINE=95129851; PubMed=782867;  
 RA Metzger R., Brown D.P., Grealish P., Staver M.J., Versalovic J.,  
 RA Lupski J.R., Katz L.;  
 RT "Characterization of the macromolecular synthesis (mms) operon from  
 Listeria monocytogenes.";  
 RL Gene 151:161-166 (1994).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=EBD-e / Serovar 1/2a;  
 RX MEDLINE=21537279; PubMed=11679669;  
 RA Glaser P., Frangeul L., Buchrieser C., Rusniok C., Amend A.,  
 RA Baquero F., Berche P., Bloeker H., Brandt P., Chakraborty T.,  
 RA Charbit A., Cherouani F., Couve E., de Darvar A., Dehoux P.,  
 RA Domann E., Dominguez-Bernal G., Duchaud E., Durant L., Dussurgeat O.,  
 RA Entian K.-D., Feibi H., Garcia-del Portillo F., Garrido P.,  
 RA Gautier L., Goebel W., Gomez-Lopez N., Hain T., Hauf J., Jackson D.,  
 RA Jones L.-M., Kaerst U., Kreft J., Kuhn M., Kunst F., Kutarpak G.,  
 RA Madueno E., Maitounam A., Mata Vicente J., Ng E., Nedjari H.,  
 RA Nardoslek G., Novella S., de Pablo B., Perez-Diaz J.-C., Purcell R.,  
 RA Remmel B., Rose M., Schluter T., Simoes N., Tierrez A.,  
 RA Vazquez-Boland J.-A., Voss H., Wehland J., Cossart P.;  
 RT "Comparative genomics of *Listeria* species.";  
 RL Science 294:849-852 (2001).  
 CC -1- FUNCTION: DNA PRIMASE IS THE POLYMERASE THAT SYNTHESIZES SMALL  
 RNA PRIMERS FOR THE OKAZAKI FRAGMENTS ON BOTH TEMPLATE STRANDS AT  
 REPLICATION FORKS DURING CHROMOSOMAL DNA SYNTHESIS.  
 CC -1- COFACTOR: BINDS ONE ZINC ION PER MOLECULE (BY SIMILARITY).  
 CC -1- SUBUNIT: MONOMER (BY SIMILARITY).  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
 or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; U11165; AAC43305.1; -;  
 DR EMBL; AL591979; CAC9533.1; -;  
 DR HSP; Q9X4D0; 100Q.  
 DR L1stList; LMO01455; -;  
 DR InterPro: IPR002936; DNAPrim\_toprim.  
 DR InterPro: IPR002694; Znf\_CHC2.  
 DR Pfam; PF01751; Toprim; 1.  
 DR Pfam; PF01807; Zf-CHC2; 1.  
 DR ProDom; PD002988; Znf\_CHC2; 1.  
 DR SMART; SM00493; TOPRIM; 1.  
 DR SMART; SM00400; ZNF\_CHC; 1.  
 KW Transferase; DNA replication; DNA-directed RNA polymerase; Primosome;  
 KW Zinc-finger; Zinc; Metal-binding; Complete proteome.  
 FT ZN\_FING 39 63 CHC2-TYPE (BY SIMILARITY).  
 FT CONFLICT 57 57 I -> W (IN REF. 1).  
 FT CONFLICT 174 174 S -> A (IN REF. 1).

```

FT CONFLICT 188 188 A -> R (IN REF. 1).
FT CONFLICT 382 382 K -> R (IN REF. 1).
FT CONFLICT 391 391 E -> D (IN REF. 1).
FT CONFLICT 394 394 G -> S (IN REF. 1).
FT CONFLICT 467 468 GA -> ET (IN REF. 1).
FT CONFLICT 478 478 A -> T (IN REF. 1).
FT CONFLICT 486 486 I -> T (IN REF. 1).
FT CONFLICT 520 520 E -> K (IN REF. 1).
FT CONFLICT 549 549 A -> T (IN REF. 1).
FT CONFLICT 569 569 P -> T (IN REF. 1).
FT CONFLICT 581 581 Y -> F (IN REF. 1).
FT CONFLICT 597 597 Y -> L (IN REF. 1).
FT CONFLICT 603 604 NE -> KD (IN REF. 1).
SQ SEQUENCE 626 AA; 71756 MW; F38C75AE34CA5709 CRC64;

Query Match 44.4%; Score 40; DB 1; Length 626;
Best Local Similarity 42.9%; Pred. No. 54;
Matches 6; Conservative 4; Mismatches 4; Indels 0; Gaps 0;

Qy 1 VGYGRDNDPDRPSR 14
Db 527 IGYFAEGNDADPTK 540

RESULT 16
MSB2_YEAST STANDARD; PRT; 1306 AA.
AC P32334;
DT 01-OCT-1993 (Rel. 27, Created)
DT 01-OCT-1993 (Rel. 27, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE MSB2 protein (Multicopy suppression of a budding defect 2).
GN MSB2 OR YGR014W.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=92383951; PubMed=1514328;
RA Bender A., Pringle J.R.;
RT "A Ser/Thr-rich multicopy suppressor of a cdc24 bud emergence
RT defect."
RL Yeast 8:315-323(1992).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=S288C;
RX MEDLINE=97435481; PubMed=9290212;
RA Rieger M., Brueckner M., Schaefer M., Mueller-Auer S.;
RT "Sequence analysis of 203 kilobases from Saccharomyces cerevisiae
RT chromosome VII."
RL Yeast 13:1077-1090(1997).
CC -1- PFM: O-GLYCOSYLATED IN THE SER/THR-RICH REGIONS (PROBABLE).
CC -1- SIMILARITY: SOME, TO YEAST HKR1.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M77354; AAA34798.1; -.
DR EMBL; Z72799; CAA96897.1; -.
DR PIR; S25370; S25370.
DR SGD; S0003246; MSB2.
KM Transmembrane; Glycoprotein; Repeat.
FT DOMAIN 698 816 7 X 17 AA TANDEM REPEATS.
FT REPEAT 698 714 1.
FT REPEAT 715 731 2.
FT REPEAT 732 748 3.
FT REPEAT 749 765 4.

```

```

FT REPEAT 766 782 5.
FT REPEAT 783 799 6.
FT REPEAT 800 816 7.
SQ SEQUENCE 1306 AA; 133114 MW; 67D5D984D5CA4A6D CRC64;

Query Match 44.4%; Score 40; DB 1; Length 1306;
Best Local Similarity 46.7%; Pred. No. 1.2e+02;
Matches 7; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

Qy 2 GYGRDNDPDRPSRN 16
Db 1269 GHGLTNDSTPRRN 1283

RESULT 17
ARS2_CAEEL STANDARD; PRT; 712 AA.
ID ARS2_CAEEL
AC Q966L5;
DT 15-JUN-2002 (Rel. 41, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Arsenite-resistance protein 2 homolog.
GN E01A2.2.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditidae;
OC Rhabditidae; Peioderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Bristol N2;
RA Wilson R., Bentley D., Scheet P.;
RL Submitted (MAR-1999) to the EMBL/GenBank/DBJ databases.
CC -1- SIMILARITY: BELONGS TO THE ARS2 FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AC006627; AAK85459.1; -.
DR WormPep; E01A2.2; CE20643.
KM Hypothetical protein.
SQ SEQUENCE 712 AA; 80794 MW; 54B051AADFA547DE CRC64;

Query Match 43.9%; Score 39.5; DB 1; Length 712;
Best Local Similarity 66.7%; Pred. No. 75;
Matches 8; Conservative 2; Mismatches 1; Indels 1; Gaps 1;

Qy 2 GYGRDNDPDRP 12
Db 645 GYGRERDDRP 656

RESULT 18
CYMS_RAT STANDARD; PRT; 146 AA.
ID CYMS_RAT
AC P04166; G90WG1;
DT 20-MAR-1987 (Rel. 04, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Cytochrome b5 outer mitochondrial membrane isoform precursor.
GN CYB5M OR OMB5.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RA Kuroda R., Ikenoue T., Honsho M., Tujimoto S., Miroma J., Ito A.;
RT "Charged amino acids at the carboxy-terminal portions determine

```

RT intracellular locations of two isoforms of cytochrome b5.";  
 RL Submitted (APR-1997) to the EMBL/GenBank/DBJ databases.  
 RN [2]  
 RP SEQUENCE OF 12-103.  
 RX MEDLINE=8318249; PubMed=6840088;  
 RA Lederer F., Ghir R., Guidard B., Cortial S., Ito A.;  
 RT "Two homologous cytochromes b5 in a single cell.";  
 RL Eur. J. Biochem. 132:95-102 (1983).  
 RN [3]  
 RP X-RAY CRYSTALLOGRAPHY (2.7 ANGSTROMS).  
 RX MEDLINE=97128656; PubMed=8973214;  
 RA Rodriguez-Maraton M.J., Qiu F., Stark R.E., White S.P., Zhang X.,  
 RA Foundling S.L., Rodriguez V., Schilling C.L. III, Bunce R.A.,  
 RA Rivera M.;  
 RT "13C NMR spectroscopic and X-ray crystallographic study of the role  
 RT played by mitochondrial cytochrome b5 heme propionates in the  
 RL electrostatic binding to cytochrome c.";  
 RL Biochemistry 35:16378-16390 (1996).  
 RN [4]  
 RP X-RAY CRYSTALLOGRAPHY (2.0 ANGSTROMS).  
 RX MEDLINE=98145212; PubMed=9484218;  
 RA Rivera M., Seetharaman R., Girdhar D., Wirtz M., Zhang X., Wang X.,  
 RA White S.;  
 RT "The reduction potential of cytochrome b5 is modulated by its exposed  
 RT heme edge.";  
 RL Biochemistry 37:1485-1494 (1998).  
 RN [5]  
 RP X-RAY CRYSTALLOGRAPHY (1.80 ANGSTROMS) OF 17-102.  
 RX MEDLINE=21041246; PubMed=11197480; Studer J., Rivera M.;  
 RA Wirtz M., Oganiesyan V., Zhang X.,  
 RT "Modulation of redox potential in electron transfer proteins: effects  
 RT of complex formation on the active site microenvironment of cytochrome  
 RT b5.";  
 RL Faraday Discuss. 116:221-234 (2000).  
 RN [6]  
 RP X-RAY CRYSTALLOGRAPHY (2.00 ANGSTROMS) OF 17-103.  
 RX MEDLINE=21466837; PubMed=11583146;  
 RA Altsuve A., Silchenko S., Lee K.-H., Kuczerka K., Terzyan S., Zhang X.,  
 RA Benson D.R., Rivera M.;  
 RT "Probing the differences between rat liver outer mitochondrial  
 RT membrane cytochrome b5 and microsomal cytochromes b5.";  
 RL Biochemistry 40:9469-9483 (2001).  
 CC -1- FUNCTION: CYTOCHROME B5 IS A MEMBRANE BOUND HEMOPROTEIN WHICH  
 CC FUNCTION AS AN ELECTRON CARRIER FOR SEVERAL MEMBRANE BOUND  
 CC OXYGENASES.  
 CC -1- SUBCELLULAR LOCATION: Mitochondrial outer membrane.  
 CC -1- SIMILARITY: BELONGS TO THE CYTOCHROME B5 FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sdb.ch/announce/>  
 CC or send an email to [license@isb-sdb.ch](mailto:license@isb-sdb.ch)).  
 CC -----  
 DR EMBL; Y12517; CA73117.1; -;  
 DR PIR; A00173; CBRTSM.  
 DR PDB; 1B5M; 12-MAR-97.  
 DR PDB; 1AMP; 18-NOV-98.  
 DR PDB; 1EUE; 04-APR-01.  
 DR PDB; 1ICC; 19-SEP-01.  
 DR InterPro; IPR001199; Cyt B5.  
 DR Pfam; PF00173; heme\_1; 1\_1-  
 DR PRINTS; PR00363; CYTOCHROMB5.  
 DR PRODOM; PD000612; Cyt B5; 1.  
 DR PROSITE; PS00191; CYTOCHROME B5\_1; 1.  
 DR PROSITE; PS50255; CYTOCHROME B5\_2; 1.  
 KW Electron transport; Mitochondrion; Outer membrane; Transmembrane;  
 KW Heme; Iron; 3D-structure.  
 FT PROPEP 1 11  
 FT CHAIN 12 146  
 FT  
 CYTOCHROME B5 OUTER MITOCHONDRIAL  
 MEMBRANE ISOFORM.

FT DOMAIN 12 103 HEME-BINDING.  
 FT TRANSMEM 119 136 POTENTIAL.  
 FT METAL 55 55 IRON (HEME AXIAL LIGAND).  
 FT METAL 79 79 IRON (HEME AXIAL LIGAND).  
 FT CONFLICT 12 12 N -> D (IN REF. 2).  
 SO SEQUENCE 146 AA; 16265 MW; 1CA9DDDC81C412E CRC64;  
 Query Match 43.3%; Score 39; DB 1; Length 146;  
 Best Local Similarity 60.0%; Pred. No. 16;  
 Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;  
 QY 7 DNDRDPKRN 16  
 DB 104 DGDKPSKRN 113  
 RESULT 19  
 CAV2\_HUMAN STANDARD; PRT; 162 AA.  
 ID CAV2\_HUMAN  
 AC P51636; Q9UGW7;  
 DT 01-OCT-1996 (Rel. 34, Created)  
 DT 15-JUL-1999 (Rel. 38, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE Caveolin-2.  
 GN CAV2.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
 CC NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.  
 RX MEDLINE=96133891; PubMed=8552590;  
 RA Scherer P.E., Okamoto T., Chun M., Nishimoto I., Lodish H.F.,  
 RA Lisanti M.P.;  
 RT "Identification, sequence, and expression of caveolin-2 defines a  
 RT caveolin gene family.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 93:131-135 (1996).  
 RN [2]  
 RP SEQUENCE FROM N.A., AND CHARACTERIZATION.  
 RX MEDLINE=98030620; PubMed=9361015;  
 RA Scherer P.E., Lewis R.Y., Volonte D., Engelman J.A., Galbardi F.,  
 RA Conet J., Kohrt D.S., Van Donselaar B., Peters P., Lisanti M.P.;  
 RT "Cell-type and tissue-specific expression of caveolin-2. Caveolins 1  
 RT and 2 co-localize and form a stable hetero-oligomeric complex in  
 RT vivo.";  
 RL J. Biol. Chem. 272:29337-29346 (1997).  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=99232848; PubMed=10218480;  
 RA Engelman J.A., Zhang X.L., Lisanti M.P.;  
 RT "Sequence and detailed organization of the human caveolin-1 and -2  
 RT genes located near the D7S222 locus (7q31.1). Methylation of a CpG  
 RT island in the 5' promoter region of the caveolin-1 gene in human  
 RT breast cancer cell lines.";  
 RL FEBS Lett. 448:221-230 (1999).  
 RN [4]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Kidney;  
 RA Strausberg R.;  
 RN Submitted (MAR-2001) to the EMBL/GenBank/DBJ databases.  
 RL [5]  
 RP SEQUENCE OF 1-10 FROM N.A.  
 RA Fra A.M., Pasqualetto E., Mancini M., Stia R.;  
 RT "Genomic organization of the human locus coding for caveolin-1 and  
 RT caveolin-2.";  
 RL Submitted (MAY-1999) to the EMBL/GenBank/DBJ databases.  
 CC -1- FUNCTION: MAY ACT AS A SCAFFOLDING PROTEIN WITHIN CAVEOLAR  
 CC MEMBRANES. INTERACTS DIRECTLY WITH G-PROTEIN ALPHA SUBUNITS AND  
 CC CAN FUNCTIONALLY REGULATE THEIR ACTIVITY. CAVEOLIN-2 MAY FUNCTION  
 CC AS AN ACCESSORY PROTEIN IN CONJUNCTION WITH CAVEOLIN-1.  
 CC -1- SUBUNIT: HOMODIMER. CAVEOLIN-1 AND -2 CO-LOCALIZE AND FORM A  
 CC STABLE HETERO-OLIGOMERIC COMPLEX.  
 CC -1- SUBCELLULAR LOCATION: MEMBRANE PROTEIN OF CAVEOLAE. POTENTIAL

```

CC HAIRPIN-LIKE STRUCTURE IN THE MEMBRANE (BY SIMILARITY).
CC -1- ALTERNATIVE PRODUCTS: BY USING ALTERNATIVE CODONS IN
CC THE SAME READING FRAME, TWO ISOPRIMS (ALPHA- AND BETA-CAVEOLIN)
CC ARE PRODUCED.
CC -1- TISSUE SPECIFICITY: EXPRESSED IN ENDOTHELIAL CELLS, SMOOTH MUSCLE
CC CELLS, KELETAL MYOBLASTS AND FIBROBLASTS.
CC -1- SIMILARITY: BELONGS TO THE CAVEOLIN FAMILY.
CC -----
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AF035752; AAB88492.1; -
DR EMBL; AJ133269; CAB63653.1; -
DR EMBL; BC005256; AAH05256.1; -
DR EMBL; AJ242718; CAB65090.1; -
DR Genbank; HGNC:1528; CAV2.
DR MIM; 601048; -
DR InterPro; IPR001612; Caveolin.
DR Pfam; PF01146; Caveolin; 1.
DR PROSITE; PS01210; CAVEOLIN; 1.
KM Transmembrane; Lipoprotein; Alternative initiation; Polymorphism.
FT CHAIN 1 162 CAVEOLIN-2 ALPHA.
FT INIT 14 162 CAVEOLIN-2 BETA.
FT DOMAIN 14 14 FOR CAVEOLIN-2 BETA.
FT TRANSMEM 87 107 CYTOPLASMIC (POTENTIAL).
FT DOMAIN 108 162 POTENTIAL.
FT VARIANT 130 130 CYTOPLASMIC (POTENTIAL).
FT SEQUENCE 130 130 0 -> E (IN DBSNP:1055892).
SQ SEQUENCE 162 AA; 18291 MW; 89FDEDA861330B87 CRC64;

Query Match 43.3%; Score 39; DB 1; Length 162;
Best Local Similarity 50.0%; Pred. No. 18;
Matches 7; Conservative 1; Mismatches 6; Indels 0; Gaps 0;

QY 3 YGRDDNDPDRSRKN 16
DB 33 FADSDQDRDPHRLN 46

RESULT 20
ID YEBN_ECOLI STANDARD; PRT; 206 AA.
AC P76264; 007974;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical protein yebn.
GN YEBN OR B1821.
OS Escherichia coli.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia coli.
OC Escherichia.
OC NCBI_Taxid=562;
OX NCBI_Taxid=562;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=K12 / MG1655;
RX MEDLINE=97426617; PubMed=9278503;
RA Blatter F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA Riley M., Coliaco-Vides J., Glaesner J.D., Rode C.K., Mayhew G.F.,
RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA Mau B., Shao Y.;
RT "The complete genome sequence of Escherichia coli K-12.";
RL Science 277:1453-1474(1997).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=K12;
RX MEDLINE=97251358; PubMed=9097040;
RA Itoh T., Aiba H., Baba T., Fujita K., Hayashi K., Inada T., Isono K.,

```

```

RA Kaeai H., Kimura S., Kitakawa M., Kitagawa M., Makino K., Miki T.,
RA Mitsuuchi K., Mori H., Mori T., Motomura K., Nakade S., Nakamura Y.,
RA Nishimoto H., Nishio Y., Oshima T., Saito N., Samped G., Seki Y.,
RA Sivasubram S., Tagami H., Takeda J., Takemoto K., Wada C.,
RA Yamamoto Y., Horuchi T.;
RT "A 460-kb DNA sequence of the Escherichia coli K-12 genome
RT corresponding to the 40.1-50.0 min region on the linkage map.";
RL DNA Res. 3:379-392(1996).
CC -1- SUBCELLULAR LOCATION: Integral membrane protein (Potential).
CC -1- SIMILARITY: BELONGS TO THE UPF0059 FAMILY.
CC -----
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AB000276; AAC74891.1; -
DR EMBL; D90826; BAA15632.1; -
DR EcoGene; EG14012; yebn.
DR InterPro; IPR003810; DUF204.
DR Pfam; PF02659; DUF204; 2.
KM Hypothetical protein; Transmembrane; Complete proteome.
FT TRANSMEM 17 37 POTENTIAL.
FT TRANSMEM 59 79 POTENTIAL.
FT TRANSMEM 84 104 POTENTIAL.
FT TRANSMEM 138 158 POTENTIAL.
FT TRANSMEM 161 181 POTENTIAL.
FT TRANSMEM 186 206 POTENTIAL.
SQ SEQUENCE 206 AA; 22050 MW; EE4876897741B02F CRC64;

Query Match 43.3%; Score 39; DB 1; Length 206;
Best Local Similarity 46.2%; Pred. No. 23;
Matches 6; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

QY 4 GRDDNDPDRSRKN 16
DB 109 GADDEDESRPRRH 121

RESULT 21
ID NAB4_YEAST STANDARD; PRT; 534 AA.
AC Q99383; Q02741;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Nuclear polyadenylated RNA-binding protein NAB4.
GN NAB4 OR HRP1 OR NAB5 OR YOL133W.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycos.
OC NCBI_Taxid=4932;
OX NCBI_Taxid=4932;
RN [1]
RP SEQUENCE FROM N.A.
RA Oberdorf A.M., Anderson J.T., Devore D.R., Swanson M.S.;
RL Submitted (SEP-1995) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=S288C;
RX MEDLINE=96366411; PubMed=8770588;
RA Henry M., Borland C.Z., Bossie W., Silver P.A.;
RT "Potential RNA binding proteins in Saccharomyces cerevisiae
RT identified as suppressors of temperature-sensitive mutations in
RT NPL3.";
RL Genetics 142:103-115(1996).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=S288C / FY1679;
RX MEDLINE=97051591; PubMed=8896268;
RA Lafuente M.J., Gamo F.J., Gancedo C.;

```

```

RT "DNA sequence analysis of a 10 624 bp fragment of the left arm of
RT chromosome XV from Saccharomyces cerevisiae reveals a RNA binding
RT protein, a mitochondrial protein, two ribosomal proteins and new
RT open reading frames."
RL Yeast 12:1041-1045 (1996).
CC -1 SUBCELLULAR LOCATION: Nuclear (Potential).
CC -1 SIMILARITY: CONTAINS 2 RNA RECOGNITION MOTIFS (RRM).
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
DR EMBL; U35737; AAA79097.1; -.
DR EMBL; U36535; AAB18142.1; -.
DR EMBL; X95258; CAA64546.1; -.
DR EMBL; Z74865; CAA93142.1; -.
DR HSSP; P09651; IUP1.
DR SGD; S0005483; HRP1.
DR InterPro; IPR00504; RNA_rec_mot.
DR Pfam; PF00076; rrm; 2.
DR SMART; SM00360; RRM; 2.
DR PROSITE; PS50102; RRM; 2.
DR PROSITE; PS00030; RRM_RNP_1; 2.
KW RNA-binding; Nuclear protein; Repeat.
FT DOMAIN 159 241 RNA-BINDING (RRM) 1.
FT FT 243 320 RNA-BINDING (RRM) 2.
SQ SEQUENCE 534 AA; 59650 MW; 1B231069437B093D CXC64;

Query Match 43.3%; Score 39; DB 1; Length 534;
Best Local Similarity 70.0%; Pred. No. 66;
Matches 7; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 2 GYGRDNDND 11
Db 486 GYNRDGRD 495

RESULT 22
HEXA_CAEEL STANDARD; PRT; 555 AA.
AC Q22492;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Probable beta-hexosaminidase A precursor (EC 3.2.1.52) (N-acetyl-beta-
DE glucosaminidase) (Beta-N-acetylhexosaminidase).
GN T1AF9.3.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditidae;
OC Rhabditidae; Peloderiinae; Caenorhabditis.
RX NCBI_TaxID=6239;
XX [1]
XX SEQUENCE FROM N.A.
XX STRAIN=Bristol N2;
XX Miller N., Langston Y.;
XX Submitted (MAR-1996) to the EMBL/GenBank/DBJ databases.
CC -1 FUNCTION: BETA-HEXOSAMINIDASE A IS RESPONSIBLE FOR THE DEGRADATION
CC OF GM2 GANGLIOSIDES, AND A VARIETY OF OTHER MOLECULES CONTAINING
CC TERMINAL N-ACETYL HEXOSAMINES.
CC -1 CATALYTIC ACTIVITY: Hydrolysis of terminal non-reducing N-
CC acetyl-D-hexosamine residues in N-acetyl-beta-D-hexosaminides.
CC -1 SUBCELLULAR LOCATION: Lysosomal (Potential).
CC -1 SIMILARITY: BELONGS TO FAMILY 20 OF GLYCOSYL HYDROLASES.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial

```

```
CC entitle requires a license agreement (See http://www.isb-sib.ch/announce/  
or send an email to license@isb-sib.ch).  
-----  
DR EMBL; U50199; AAA91263.1; -.  
DR HSSP; P07686; IQBD.  
DR WormPeP; T14F9.3; CE07499.  
DR InterPro; IPR001540; GH_20.  
DR Pfam; PF00728; Glyco_hydro_20; 1.  
DR Pfam; PF00838; Glyco_hydro_20; 1.  
DR PRINTS; PR00738; GLHYDRLASE20.  
KW Hydroxylase; Glycosidase; Lysosome; Signal; Glycoprotein.  
FT SIGNAL 1 18  
FT CHAIN 19 555 POTENTIAL.  
FT CARBOHYD 47 47 PROBABLE BETA-HEXOSAMINIDASE A.  
FT CARBOHYD 351 351 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 412 412 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 460 460 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT ACT_SITE 325 325 CATALYTIC ACID (BINDS TO THE GLYCOSIDIC LINKAGE) (BY SIMILARITY).  
SQ SEQUENCE 555 AA; 64379 MW; 5E797FP9D826B84 CRC64;  
  
Oy 3 YGRDPNDR 10  
Db 21 YGRDDPDR 28  
  
RESULT 23  
YSPK CAEEL STANDARD; PRT; 615 AA.  
ID Q19425;  
AC 01-JUN-1997 (Rel. 35, Created)  
DT 15-JUN-2002 (Rel. 41, Last sequence update)  
DT 15-JUN-2002 (Rel. 41, Last annotation update)  
DE Hypothetical protein F13H10.3 in chromosome IV.  
GN F13H10.3.  
OS Caenorhabditis elegans.  
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditiida; Rhabditoidea;  
OC Rhabditiidae; Peloderinae; Caenorhabditis.  
OX NCBI_TaxID=6239;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=Bristol N2;  
RA Cottage A.;  
RL Submitted (JAN-1996) to the EMBL/GenBank/DDBJ databases.  
RN [2]  
RP REVISIONS.  
RA Durbin R.;  
RL Submitted (OCT-2001) to the EMBL/GenBank/DDBJ databases.  
CC -1- SUBCELLULAR LOCATION: Integral membrane protein (Potential).  
CC -1- SIMILARITY: SOME, TO YEAST YEL064C.  
-----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/  
or send an email to license@isb-sib.ch).  
-----  
DR EMBL; Z68748; CAA92953.2; -.  
DR WormPeP; F13H10.3; CE27959.  
DR InterPro; IPR002422; AA/rej_primease2.  
DR Pfam; PF01490; Aa_trans; 1.  
KW Hypothetical protein; Transmembrane.  
FT TRANSMEM 166 186 POTENTIAL.  
FT TRANSMEM 195 215 POTENTIAL.  
FT TRANSMEM 248 268 POTENTIAL.  
FT TRANSMEM 342 362 POTENTIAL.  
FT TRANSMEM 368 388 POTENTIAL.
```



```
FT TRANSMEM 416 436 POTENTIAL.
FT TRANSMEM 452 472 POTENTIAL.
FT TRANSMEM 496 516 POTENTIAL.
FT TRANSMEM 536 556 POTENTIAL.
FT TRANSMEM 562 582 POTENTIAL.
FT TRANSMEM 595 615 POTENTIAL.
SQ SEQUENCE 615 AA; 68692 MW; 5092F2ABF4A3A420 CRC64;

Query Match 43.3%; Score 39; DB 1; Length 615;
Best Local Similarity 77.8%; Pred. No. 77;
Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 4 GRDNDNDP 12
Db 40 GVDNDTDP 48

RESULT 24
ID DDIS_CAEEL STANDARD; PRT; 739 AA.
AC Q20875;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Putative pre-mRNA splicing factor ATP-dependent RNA helicase F56D2.6.
GN F56D2.6.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditioidea;
OC Rhabditidae; Pelodetinae; Caenorhabditis.
OX NCBI_TaxId=6239;
RN
RA Du 2;
RC STRAIN=Bristol N2;
RL Submitted (NOV-1994) to the EMBL/GenBank/DDBJ databases.
CC -!- FUNCTION: PRE-MRNA PROCESSING FACTOR INVOLVED IN DISASSEMBLY OF
CC SPLICEOSOMES AFTER THE RELEASE OF MATURE MNNA (BY SIMILARITY).
CC -!- SUBCELLULAR LOCATION: Nuclear (By similarity).
CC -!- SIMILARITY: BELONGS TO THE DEAD BOX HELICASE FAMILY. DEAH
CC SUBFAMILY. DDx15/PRP43 ORTHOLOG.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; U13644; AAB52678.1; -.
DR WormPep; F56D2.6; CE01334.
DR InterPro; IPR001410; DEAD.
DR InterPro; IPR002464; DEAH_box.
DR InterPro; IPR001650; Helicase_C.
DR Pfam; PF00271; helicase_C; 1.
DR SMART; SM00487; DEXDC; 1.
DR SMART; SM00490; HELIC_C; 1.
DR PROSITE; PS00690; DEAH_ATP_HELICASE; FALSE_NEG.
KW Hypothetical protein; mRNA processing; mRNA splicing; Helicase;
KW ATP-binding; Nuclear protein.
FT NP_BIND 99 106 ATP (POTENTIAL).
FT DOMAIN 119 124 POLY-GLN.
FT SITE 204 207 DEAH BOX.
SQ SEQUENCE 739 AA; 84388 MW; AB9FBDCFE22F40F CRC64;

Query Match 43.3%; Score 39; DB 1; Length 739;
Best Local Similarity 46.7%; Pred. No. 95;
Matches 7; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

Qy 2 GYGRDNDNDPDRPSRKV 16
Db 11 GSGRGDRRRSPNRRS 25
```

```
RESULT 25
ID VGLN_CHICK STANDARD; PRT; 1270 AA.
AC P81021;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Vigilin.
GN HDLBP OR VGL.
OS Gallus gallus (chicken).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC Gallus.
OX NCBI_TaxId=9031;
RN
RA [1]
RC SEQUENCE FROM N.A.
RC TISSUE=Embryonic sternal cartilage;
RX MEDLINE=92298988; PubMed=1606952;
RA Schmidt C., Henkel B., Poeschl E., Zorbas H., Purschke W.G.,
RA Gloe T.R., Mueller P.K.;
RT "Complete cDNA sequence of chicken vigilin, a novel protein with
RT amplified and evolutionary conserved domains.";
RL Eur. J. Biochem. 206:625-634(1992).
CC -!- SUBCELLULAR LOCATION: INTRACELLULAR.
CC -!- SIMILARITY: CONTAINS 14 KH DOMAINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X65292; CA46387.1; -.
DR InterPro; IPR004087; KH_dom.
DR InterPro; IPR004088; KH_type_1.
DR Pfam; PF00013; KH-domain; 14.
DR SMART; SM00322; KH; 14.
DR PROSITE; PS50084; KH_TYPE_1; 14.
KW RNA-binding; Repeat.
FT DOMAIN 150 188 KH 1.
FT DOMAIN 199 260 KH 2.
FT DOMAIN 291 333 KH 3.
FT DOMAIN 360 402 KH 4.
FT DOMAIN 431 473 KH 5.
FT DOMAIN 504 545 KH 6.
FT DOMAIN 577 619 KH 7.
FT DOMAIN 651 693 KH 8.
FT DOMAIN 724 766 KH 9.
FT DOMAIN 798 840 KH 10.
FT DOMAIN 872 913 KH 11.
FT DOMAIN 970 1012 KH 12.
FT DOMAIN 1051 1093 KH 13.
FT DOMAIN 1126 1168 KH 14.
SQ SEQUENCE 1270 AA; 142220 MW; 0A8844F91F9B8619 CRC64;

Query Match 43.3%; Score 39; DB 1; Length 1270;
Best Local Similarity 58.3%; Pred. No. 1,7e+02;
Matches 7; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

Qy 2 GYGRDNDNDPDS 13
Db 933 GEGKDGKADPS 944

RESULT 26
ID SSM4_YEAST STANDARD; PRT; 1319 AA.
AC P40318;
DT 01-FEB-1995 (Rel. 31, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
```

DT 01-NOV-1997 (Rel. 35, last annotation update)  
 DE SSM4 protein.  
 GN SSM4 OR Y1030C OR Y13299.01C OR Y19905.18C.  
 OS Saccharomyces cerevisiae (Baker's yeast).  
 CC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
 OC Saccharomycetales; Saccharomycetaceae; Saccharomyces.  
 ON NCBI\_TaxID=4932;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=ATCC 28363 / FL100;  
 RX MEDLINE=95115683; PubMed=7816042;  
 RA Mandart E., Dufour M.-E., Lacroix F.;  
 RT "Inactivation of SSM4, a new Saccharomyces cerevisiae gene,  
 suppresses mRNA instability due to RNA14 mutations.";  
 RL Mol. Gen. Genet. 245:323-333(1994).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=S288c / AB972;  
 RA Barrell B.G., Badcock K., Bankier A.T., Bowman S., Brown D.,  
 RA Churcher C.M., Connor R., Copsey T., Dear S., Devlin K., Fraser A.,  
 RA Goules S., Hamlyn N., Horenell T.S., Hunt S., Jagsis K., Jones M.,  
 RA Louis E., Lyne G., Moul S., Moul T., Odell C., Pearson D.,  
 RA Rajandream M.A., Riles L., Rowley N., Skelton J., Smith V.,  
 RA Walsh S.V., Whitehead S.;  
 RL Submitted (DEC-1994) to the EMBL/GenBank/DBJ databases.  
 RN [3]  
 RP SEQUENCE OF 1-392 FROM N.A.  
 RC STRAIN=S288c / AB972;  
 RA Skelton J., Bowman S., Churcher C., Barrell B.G., Rajandream M.A.;  
 RL Submitted (DEC-1994) to the EMBL/GenBank/DBJ databases.  
 RN [4]  
 RP SEQUENCE OF 383-1319 FROM N.A.  
 RC STRAIN=S288c / AB972;  
 RL Submitted (DEC-1994) to the EMBL/GenBank/DBJ databases.  
 CC -1- FUNCTION: MAY BE A NUCLEASE WHICH WOULD NORMALLY BE INACTIVE OWING  
 TO INHIBITION BY WILD-TYPE RNA14 PROTEIN AND COULD BE ACTIVATED  
 EITHER AS A NORMAL COMPONENT OF MRNA DECAY OR DURING SOME  
 SPECIFIC CELLULAR STRESS. IT COULD ALSO BE AN ACTIVATOR OF 3'-5'  
 RIBONUCLEASE OR MIGHT BE DIRECTLY IMPLICATED IN ASSOCIATION WITH  
 THE RNA14 PROTEIN IN THE NUCLEAR POLYADENYLATION OF THE MRNAS.  
 CC  
 CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC EMBL; X76715; CAAS4133.1; -;  
 DR EMBL; Z46881; CAAB6961.1; -;  
 DR EMBL; Z46861; CAAB6921.1; -;  
 DR SGD; S0001292; SSM4.  
 DR InterPro; IPR001202; WW\_RAP5\_WWP.  
 DR PROSITE; PS01159; WW\_DOMAIN\_1; UNKNOWN\_1.  
 FT DOMAIN 288 291 POLY-ALA.  
 FT DOMAIN 293 296 POLY-ASN.  
 FT DOMAIN 332 341 ASP/GLN/SER-RICH (ACIDIC).  
 FT DOMAIN 369 374 POLY-GLN.  
 FT DOMAIN 421 424 L->F (IN REF. 1).  
 FT CONFLICT 743 743 A->T (IN REF. 1).  
 FT CONFLICT 1085 1085 N->D (IN REF. 1).  
 FT CONFLICT 1186 1186 Y->F (IN REF. 1).  
 SO SEQUENCE 1319 AA; 151454 MW; 3EDFF7F9D90A0C8C CRC64;

Query Match 43.3%; Score 39; DB 1; Length 1319;  
 Best Local Similarity 42.9%; Pred. No. 1.8e+02;  
 Matches 6; Conservative 5; Mismatches 3; Indels 0; Gaps 0;

OY 2 GYGRDNDPDRSK 15  
 DB 598 GYGRDNDPDRSK 611

RESULT 27  
 CPXO\_SACER STANDARD; PRT; 397 AA.  
 AC P4635; O33990;  
 DT 01-FEB-1996 (Rel. 33, Created)  
 DT 15-DEC-1998 (Rel. 37, Last sequence update)  
 DT 15-DEC-1998 (Rel. 37, Last annotation update)  
 DE Cytochrome P450 113A1 (Erythromycin B/D C-12 hydroxylase).  
 GN ERYK OR CYP113A1.  
 OS Saccharopolyspora erythraea (Streptomyces erythraeus).  
 OC Bacteria; Actinobacteria; Actinobacteriales; Actinobacteridae;  
 OC Actinomycetales; Pseudonocardiales; Pseudonocardaceae;  
 OC Saccharopolyspora.  
 ON NCBI\_TaxID=1836;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=NRRL 2338;  
 RX MEDLINE=93106953; PubMed=8416893;  
 RA Stasell D., Donadio S., Staver M.J., Katz L.;  
 RT "Identification of a Saccharopolyspora erythraea gene required for  
 the final hydroxylation step in erythromycin biosynthesis.";  
 RL J. Bacteriol. 175:182-189(1993).  
 RN [2]  
 RP REVISIONS.  
 RC STRAIN=NRRL 2338;  
 RX MEDLINE=97390129; PubMed=9249068;  
 RA Pereda A., Summers R., Katz L.;  
 RT "Nucleotide sequence of the erme distal flank of the erythromycin  
 biosynthesis cluster in Saccharopolyspora erythraea.";  
 RL Gene 193:65-71(1997).  
 CC -1- FUNCTION: RESPONSIBLE FOR THE C-12 HYDROXYLATION OF THE  
 MACROLACTONE RING OF ERYTHROMYCIN.  
 CC  
 CC -1- PATHWAY: FINAL HYDROXYLATION STEP OF ERYTHROMYCIN B AND D  
 SYNTHESIS.  
 CC  
 CC -1- SIMILARITY: BELONGS TO THE CYTOCHROME P450 FAMILY.  
 CC  
 CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC EMBL; U82823; AAC45584.1; -;  
 DR PIR; B40634; B40634.  
 DR HSSP; Q55080; 1107.  
 DR InterPro; IPR001128; Cytochrome\_P450.  
 DR Pfam; PF00067; P450; 1.  
 DR PROSITE; PS00086; CYTOCHROME\_P450; 1.  
 KM Oxidoreductase; Monooxygenase; Electron transport; Heme;  
 KM Antibiotic biosynthesis.  
 FT BINDING 339 339 HEME (BY SIMILARITY).  
 SO SEQUENCE 397 AA; 43759 MW; 8306AF79414C881E CRC64;

Query Match 42.8%; Score 38.5; DB 1; Length 397;  
 Best Local Similarity 75.0%; Pred. No. 58;  
 Matches 9; Conservative 1; Mismatches 1; Indels 1; Gaps 1;

OY 6 DDNDR-DPSRKN 16  
 DB 313 DDNDR-DPSRKS 324

RESULT 28  
 YL53\_CAEEL STANDARD; PRT; 244 AA.  
 ID YL53\_CAEEL  
 AC P34433;  
 DT 01-FEB-1994 (Rel. 28, Created)  
 DT 01-FEB-1994 (Rel. 28, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)

```

DE Hypothetical protein F44E2.3 in chromosome III.
GN F44E2.3.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidae;
OC Rhabditidae; Pelodermine; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=Briscot N2;
RC MEDLINE=94150718; PubMed=7906398;
RA Wilson R., Ainscough R., Anderson K., Baynes C., Berke M., Coulson A.,
RA Bonfield W., Burton J., Connell M., Copsey T., Cooper J., Fraser A.,
RA Craxton M., Dear S., Du Z., Durbin R., Favello A., Fraser A.,
RA Fulton L., Gardner A., Green P., Hawkins T., Hillier L., Jier M.,
RA Johnston L., Jones M., Kerhaw J., Kirsten J., Laister N.,
RA Latreille P., Lightning J., Lloyd C., Mortimore B., O'Callaghan M.,
RA Parsons J., Percy C., Rifken L., Roopra A., Saunders D., Shownkeen R.,
RA Sims M., Smaildon N., Smith A., Smith M., Sonnenhammer E., Staden R.,
RA Sulston J., Thierry-Mieg J., Thomas K., Vaubin M., Vaughan K.,
RA Waterston R., Watson A., Weinstock L., Wilkinson-Sproat J.,
RA Wohldman P.;
RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
RT elegans.";
RL Nature 368:32-38(1994).
CC -1- SIMILARITY: TO DNAB.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; L23646; AAA28041.1; -.
DR PIR; S44822; S44822.
DR Wormpep; F44E2.3; CE00181.
KM Hypothetical protein.
FT DOMAIN 3 45 ARG/ASP/LYS-RICH.
FT DOMAIN 79 90 PRO-RICH.
SQ SEQUENCE 244 AA; 28994 MW; COCE677FB01A2B18 CRC64;
Query Match 42.2%; Score 38; DB 1; Length 244;
Best Local Similarity 63.6%; Pred. No. 40;
Matches 7; Conservative 1; Mismatches 3; Indels 0; Gaps 0;
QY 5 RDDNDRDPSRK 15
DB 22 REDRDREK 32
RESULT 29
PITX1_XENLA STANDARD; PRT; 305 AA.
AC Q9W751; Q9DEN6; Q9I898;
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Pituitary homeobox 1 (X-PITX-1) (Xp1tx1).
GN PITX1.
OS Xenopus laevis (African clawed frog).
OC Eukaryota; Metazoa; Chordata; Craniata; Euteleostomi;
OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipidoidea; Pipidae;
OC Xenopodinae; Xenopus.
OX NCBI_TaxID=8355;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20005803; PubMed=10534625;
RA Hollemann T., Pieler T.;
RT "Pitx-1: a homeobox gene expressed during pituitary and cement gland
RT formation of Xenopus embryos.";
RL Mech. Dev. 88:249-252(1999).

```

```

RP SEQUENCE FROM N.A.
RX MEDLINE=21113427; PubMed=1170348;
RA Chang W.Y., Khosrowshahian F., Chang R., Crawford M.J.;
RT "Pitx1 plays a role in specifying cement gland and head during early
RT Xenopus development.";
RL Genesis 29:78-90(2001).
RN [3]
RP SEQUENCE FROM N.A.
RX MEDLINE=21369728; PubMed=11477694;
RA Schweickert A., Deisler K., Blum M., Streinbeisser H.;
RT "Pitx1 and Pitx2c are required for ectopic cement gland formation in
RT Xenopus laevis.";
RL Genesis 30:144-148(2001).
CC -1- FUNCTION: May play a role in the development of anterior
CC structures, and in particular, the brain and facies and in
CC specifying the identity or structure of hindlimb.
CC -1- SUBCELLULAR LOCATION: Nuclear.
CC -1- DEVELOPMENTAL STAGE: Expressed in the anterior neural ridge and in
CC the cement gland Anlage during late gastrulation/early
CC neurulation.
CC -1- SIMILARITY: BELONGS TO THE PAIRED HOMEBOX FAMILY.
CC "BICOID" SUPERFAMILY.
CC -1- SIMILARITY: CONTAINS 1 HOMEBOX DOMAIN.
CC -1- SIMILARITY: CONTAINS 1 OAR DOMAIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AF155206; AAB45292.1; -.
DR EMBL; AF217647; AAF29531.1; -.
DR EMBL; AJ278330; CAC12834.1; -.
DR HSSP; P06601; 1FJL.
DR InterPro; IPR003654; Homeo OAR.
DR InterPro; IPR001356; Homeobox.
DR InterPro; IPR000327; POU domain.
DR Pfam; PF00046; homeobox; 1.
DR PRINTS; PR00024; HOMEBOX.
DR PRINTS; PR00028; POUDOMAIN.
DR ProDom; PD000010; Homeobox; 1.
DR SMART; SM00389; HOX; 1.
DR PROSITE; PS00027; HOMEBOX_1; 1.
DR PROSITE; PS00071; HOMEBOX_2; 1.
DR PROSITE; PS50803; OAR; 1.
KW Homeobox; DNA-binding; Developmental protein; Nuclear protein;
KW Transcription regulation; Activator.
FT DNA BIND 78 137 HOMEBOX.
FT DOMAIN 138 270 INTERACTS WITH PIT-1 (BY SIMILARITY).
FT DOMAIN 271 284 OAR.
FT DOMAIN 284 288 NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).
FT CONFLICT 7 7 G -> A (IN REF. 2 AND 3).
FT CONFLICT 54 54 A -> V (IN REF. 3).
SQ SEQUENCE 305 AA; 34128 MW; ACA215A5BA86B43F CRC64;
Query Match 42.2%; Score 38; DB 1; Length 305;
Best Local Similarity 58.3%; Pred. No. 52;
Matches 7; Conservative 1; Mismatches 4; Indels 0; Gaps 0;
QY 4 GRDNDRPSRK 15
DB 64 GEDGNDPSRK 75
RESULT 30
CATA_ACTCA STANDARD; PRT; 311 AA.
AC P07773;
DT 01-JAN-1990 (Rel. 13, Created)
DT 01-JAN-1990 (Rel. 13, Last sequence update)

```

```

DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Catechol 1,2-dioxygenase (EC 1.13.11.1).
GN CATA.
OS Acinetobacter calcoaceticus.
OC Bacteria; Proteobacteria; gamma subdivision; Moraxellaceae;
OC Acinetobacter.
ON NCBI_TaxID=471;
RX [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BD413 / ADP1;
RX MEDLINE=89008110; PubMed=3170486;
RA "DNA sequence of the Acinetobacter calcoaceticus catechol 1,2-
RT dioxygenase I structural gene cata: evidence for evolutionary
RT divergence of intradiol dioxygenases by acquisition of DNA sequence
RT repetitions."
RL J. Bacteriol. 170:4874-4880(1988).
CC -1- CATALYTIC ACTIVITY: Catechol + O(2) = cis,cis-muconate.
CC -1- COPACTOR: FERRIC ION.
CC -1- PATHWAY: FIRST STEP IN THE DEGRADATION OF CATECHOL TO SUCCINATE
CC AND ACETYL-COA IN THE BETA-KETOADIPATE PATHWAY.
CC -1- SIMILARITY: BELONGS TO THE INTRADIOL RING-CLEAVAGE DIOXYGENASE
CC FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AF009224; AAC46426.1; -.
DR HSSP; P00437; 3PCC.
DR InterPro; IPR000627; Dioxigenase.
DR Pfam; PF00775; Dioxigenase; 1.
DR PROSITE; PS00083; INTRADIOL_DIOXYGENAS; 1.
RW Aromatic hydrocarbons catabolism; Oxidoreductase; Dioxigenase; Iron.
FT METAL 164 164 IRON (BY SIMILARITY).
FT METAL 200 200 IRON (BY SIMILARITY).
FT METAL 224 224 IRON (BY SIMILARITY).
FT METAL 226 226 IRON (BY SIMILARITY).
SQ SEQUENCE 311 AA; 34347 MW; 8999FPD3F783B4DB CRC64;

Query Match 42.2%; Score 38; DB 1; Length 311;
Best Local Similarity 53.8%; Pred. No. 53;
Matches 7; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

QY 1 VGYGRDNDNDPS 13
Db 118 VGYARMDDSDPN 130

RESULT 31
DJB4_HUMAN STANDARD; PRT; 337 AA.
ID DJB4_HUMAN Q9UDY4; Q13431;
AC Q9UDY4; Q13431;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE DnaJ homolog subfamily B member 4 (Heat shock 40 kDa protein 1
DE homolog) (Heat shock protein 40 homolog) (HSP40 homolog).
GN DNAJB4 OR HJ41 OR DNAJW.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxID=9606;
RX [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Liver;
RX MEDLINE=98207686; PubMed=9546042;
RA Hoe K.L., Won M., Chung K.S., Jang Y.J., Lee S.B., Kim D.U., Lee J.W.,
RA Yun J.H., Yoo H.S.;

```

```

RT "Isolation of a new member of DnaJ-like heat shock protein 40 (Hsp40)
RT from human liver."
RT Biochim. Biophys. Acta 1383:4-8(1998).
RN [2]
RP PRELIMINARY SEQUENCE FROM N.A.
RC TISSUE=Liver;
RA Won M., Moon K.M., Lee C.E., Yoo H.S.;
RL Submitted (NOV-1995) to the EMBL/GenBank/DBJ databases.
CC -1- INDUCTION: BY HEAT SHOCK.
CC -1- SIMILARITY: CONTAINS 1 J DOMAIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; U40992; AAC14483.2; -.
DR EMBL; U41290; AAB07346.1; ALT_FRAME.
DR HSSP; P25685; 1HDJ.
DR Genew; HGNC:14886; DNAJB4.
DR InterPro; IPR002939; DnaJ_C.
DR InterPro; IPR001623; DnaJ_N.
DR InterPro; IPR003095; Hep_DnaJ.
DR Pfam; PF00226; DnaJ; 1.
DR Pfam; PF01556; DnaJ_C; 1.
DR PRINTS; PR00625; DNAJPROTEIN.
DR SMART; SM00271; DnaJ; 1.
DR PROSITE; PS00636; DnaJ_1; 1.
DR PROSITE; PS0076; DnaJ_2; 1.
KW Chaperone; Heat shock.
FT DOMAIN 1 70 J-DOMAIN.
SQ SEQUENCE 337 AA; 37806 MW; C7A9C613F73BCDAC CRC64;

Query Match 42.2%; Score 38; DB 1; Length 337;
Best Local Similarity 61.5%; Pred. No. 58;
Matches 8; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 2 GYGRDNDNDPSR 14
Db 141 GYGRDNDNDPSR 153

RESULT 32
DJB4_MOUSE STANDARD; PRT; 337 AA.
ID DJB4_MOUSE Q9D832; Q9D902;
AC Q9D832; Q9D902;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE DnaJ homolog subfamily B member 4.
GN DnaJB4.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RX [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Small intestine, and Testis;
RX MEDLINE=21085660; PubMed=11217851;
RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA Arakawa T., Hara A., Fukunishi Y., Kono H., Adachi J., Fukuda S.,
RA Aizawa K., Iwata M., Nishi K., Kiyosawa H., Kondo S., Yamanka I.,
RA Saito T., Okazaki Y., Gojopori T., Bono H., Kaikawa T., Saito R.,
RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
RA Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,
RA Schriml L.M., Strabli F., Suzuki R., Tomita M., Wagner L., Washio T.,
RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
RA Blake J., Botelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,

```

RA Guecinić S., Hill D., Hofmann M., Hume D.A., Kamlya M., Lee N.H.,  
 RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Momberts P.,  
 RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,  
 RA Saeki H., Sato K., Schoenbach C., Seta T., Shibata Y., Storch K.-P.,  
 RA Suzuki H., Toyooka K., Wang K.H., Weitz C., Whitaker C., Wilmig L.,  
 RA Wyszew-Boris A., Yoshida K., Haegawa Y., Kawaji H., Kotsuki S.,  
 RA Hayashizaki Y.,  
 RT "Functional annotation of a full-length mouse cDNA collection";  
 RL Nature 409:685-690(2001).  
 CC -1- SIMILARITY: CONTAINS 1 J DOMAIN.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL: AK008537; BAB25729.1; -  
 DR EMBL: AK006478; BAB24608.1; -  
 DR HSSP: P25685; 1HDJ.  
 DR MGD: MGI:1917260; 1700029A20R1k.  
 DR InterPro: IPR002939; DnaJ\_C.  
 DR InterPro: IPR001623; DnaJ\_N.  
 DR InterPro: IPR003095; Hsp\_DnaJ.  
 DR Pfam: PF00226; DnaJ; 1.  
 DR Pfam: PF01556; DnaJ\_C; 1.  
 DR PRINTS: PR00625; DnaJPROTEIN.  
 DR SMART: SM00271; DnaJ; 1.  
 DR PROSITE: PS00636; DnaJ\_1; 1.  
 DR PROSITE: PS50076; DnaJ\_2; 1.  
 KM Chapterone.  
 FT DOMAIN 4 68 J-DOMAIN.  
 FT CONFLICT 66 Q -> L (IN REF. 1; BAB24608).  
 FT CONFLICT 151 P -> T (IN REF. 1; BAB24608).  
 FT CONFLICT 329 I -> S (IN REF. 1; BAB24608).  
 SQ SEQUENCE 337 AA; 37782 MW; BEE4A0E25BCEFF4 CRC64;  
 QY 2 GYGRDNDNDPSR 14  
 Db 141 GYPRDNRNVGPR 153  
 Query March 42.2%; Score 38; DB 1; Length 337;  
 Best Local Similarity 61.5%; Pred. NO. 58;  
 Matches 8; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

RX MEDLINE=92388176; PubMed=1517240;  
 RA Ulysbeck M., Lazure C., Lilja H., Spurr N.K., Rao V.V., Loeffler C.,  
 RA Hansman I., Lundwall A.;  
 RT "Gene structure of semenogelin I and II. The predominant proteins in  
 RT human semen are encoded by two homologous genes on chromosome 20.";  
 RL J. Biol. Chem. 267:18080-18084(1992).  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=21638749; PubMed=11780052;  
 RA Deloukas P., Matthews L.H., Aghurst J., Burton J., Gilbert J.G.R.,  
 RA Jones M., Stavrides G., Almeida J.P., Babbage A.K., Bagguley C.L.,  
 RA Bailey J., Barlow K.F., Bates K.N., Beard L.M., Beare D.M.,  
 RA Beasley O.P., Bird C.P., Blakey S.E., Bridgeman A.M., Brown A.J.,  
 RA Buck D., Burrill W.D., Butler A.P., Carder C., Carter N.P.,  
 RA Chapman J.C., Clamp M., Clark G., Clark L.N., Clark S.Y., Clee C.M.,  
 RA Cligg S., Cobley V.E., Collier R.E., Connor R.E., Corby N.R.,  
 RA Coulson A., Coville G.J., Deadman R., Dhani P.D., Dunn M.,  
 RA Ellington A.G., Frankland J.A., Fraser A., French L., Garner P.,  
 RA Graffham D.V., Griffiths C., Griffiths M.N.D., Gwilliam R., Hall R.E.,  
 RA Hammond S., Harley J.L., Heath P.D., Ho S., Holden J.L., Howden P.J.,  
 RA Huckle E., Hunt A.R., Hunt S.E., Jekosch K., Johnson C.M., Johnson D.,  
 RA Kay M.P., Kimberley A.M., King A., Knights A., Laird G.K., Lawlor S.,  
 RA Leharalain M.H., Levertha M.A., Lloyd C., Lloyd D.M., Lovell J.D.,  
 RA Marsh V.L., Martin S.L., McConnachie L.J., McIay J.C., McMurtry A.A.,  
 RA Milne S.A., Mistry D., Moore M.J.F., Mullikin J.C., Nickerson T.,  
 RA Oliver K., Parker A., Patel R., Pearce T.A.V., Peck A.I.,  
 RA Phillimore B.J.C.T., Prathalingam S.R., Plumb R.W., Ramsay H.,  
 RA Rice C.M., Ross M.T., Scott C.E., Sehra H.K., Shownkeen R., Sims S.,  
 RA Skuce C.D., Smith M.L., Soderlund C., Steward C.A., Sulston J.E.,  
 RA Swann R.M., Sycamore N., Taylor R., Tee L., Thomas D.W., Thorpe A.,  
 RA Tracey A., Tromans A.C., Vaudin M., Wall M., Wallis J.M.,  
 RA Whitehead S.L., Whitaker P., Willey D.L., Williams L., Williams S.A.,  
 RA Wilmig L., Wray P.W., Hubbard T., Durbin R.M., Bentley D.R., Beck S.,  
 RA Rogers J.;  
 RT "The DNA sequence and comparative analysis of human chromosome 20.";  
 RL Nature 414:865-871(2001).  
 RN [4]  
 RP SEQUENCE OF 108-159.  
 RX MEDLINE=85127550; PubMed=3972122;  
 RA Lilja H., Jeppsson J.-O.;  
 RT "Partial amino acid sequence of a human seminal plasma peptide with  
 RT inhibin-like activity.";  
 RL FEBS Lett. 182:181-184(1985).  
 RN [5]  
 RP SEQUENCE OF 108-138.  
 RX MEDLINE=84132557; PubMed=6698208;  
 RA Seidah N.G., Ramasarma K., Saitam M.R., Chretien M.;  
 RT "Partial amino acid sequence of a human seminal plasma peptide with  
 RT inhibin-like activity.";  
 RL FEBS Lett. 167:98-102(1984).  
 RN [6]  
 RP SEQUENCE OF 68-159.  
 RX MEDLINE=85216629; PubMed=3889920;  
 RA Li C.H., Hammonds R.G., Ramasarma K., Chung D.;  
 RT "Human seminal alpha inhibins: isolation, characterization, and  
 RT structure.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 82:4041-4044(1985).  
 CC -1- FUNCTION: SEMENOGELIN I IS THE PREDOMINANT PROTEIN IN HUMAN SEMEN.  
 CC IT PARTICIPATES IN THE FORMATION OF A GEL MATRIX ENTRAPPING THE  
 CC ACCESSORY GLAND SECRETIONS AND EJACULATED SPERMATOZOA. FRAGMENTS  
 CC OF SEMENOGELIN AND/OR FRAGMENTS OF THE RELATED PROTEINS MAY  
 CC CONTRIBUTE TO THE ACTIVATION OF PROGRESSIVE SPERM MOVEMENTS AS  
 CC THE GEL-FORMING PROTEINS ARE FRAGMENTED BY THE KALLIKREIN-LIKE  
 CC PROTEASE.  
 CC -1- FUNCTION: ALPHA-INHIBIN-92 AND ALPHA-INHIBIN-31, DERIVED FROM  
 CC THE PROTEOLYTIC DEGRADATION OF SEMENOGELIN, INHIBIT THE SECRETION  
 CC OF PITUITARY FOLLICLE-STIMULATING HORMONE.  
 CC -1- SUBUNIT: OCCURS IN DISULFIDE-LINKED COMPLEXES WHICH MAY ALSO  
 CC CONTAIN TWO LESS ABUNDANT 71- AND 76-KDA SEMENOGELIN-RELATED  
 CC POLYPEPTIDES.  
 CC -1- TISSUE SPECIFICITY: SEMINAL VESICLE.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.ebi.ac.uk/announcement/>  
 CC or send an email to [license@ebi.ac.uk](mailto:license@ebi.ac.uk)).

CC -----  
 DR EMBL; J04440; AAB59506.1; -;  
 DR EMBL; M81650; AAB18168.1; -;  
 DR EMBL; A049767; CAB53523.1; -;  
 DR PIR; A03254; WTHUB.  
 DR PIR; A31489; A31489.  
 DR Genew; HGNC:10742; SEMG1.  
 DR MIM; 182140; -;  
 DR Semem; Seminal vesicle; Repeat; Signal; Polymorphism.  
 FT SIGNAL 1 23  
 FT CHAIN 24 462 SEMENOGELIN I.  
 FT PEPTIDE 68 159 ALPHA-INHIBIN-92.  
 FT PEPTIDE 108 138 ALPHA-INHIBIN-31.  
 FT MOD RES 24 24 SEMINAL BASIC PROTEIN.  
 FT DISULFID 239 239 PYROLIDONE CARBOXYLIC ACID (PROBABLE).  
 FT REPEAT 174 215 INTERCHAIN.  
 FT REPEAT 234 275 42 AA REPEAT 1.  
 FT REPEAT 282 339 42 AA REPEAT 2.  
 FT REPEAT 342 399 58 AA REPEAT 1.  
 FT REPEAT 414 455 58 AA REPEAT 2.  
 FT VARIANT 79 79 42 AA REPEAT 3.  
 FT S -> T (LESS COMMON GENETIC VARIANT).  
 FT /FtId=VAR\_005610.  
 FT K -> N (IN REF. 2).  
 SQ SEQUENCE 462 AA; 52131 MM; 760F48FCF2FA702 CRC64;  
 Query Match 42.2%; Score 38; DB 1; Length 462;  
 Best Local Similarity 57.1%; Pred. No. 82;  
 Matches 8; Conservative 1; Mismatches 5; Indels 0; Gaps 0;

QY 3 YGRDNDRDPKRN 16  
 | | | | |  
 Db 69 YHVANDHDQSRKS 82

RESULT 34  
 WRK4\_ARATH STANDARD; PRT; 514 AA.  
 ID WRK4\_ARATH  
 AC O9X190; O93WN8;  
 DT 15-JUN-2002 (Rel. 41, Created)  
 DT 15-JUN-2002 (Rel. 41, Last sequence update)  
 DE Probable WRKY transcription factor 4 (WRKY DNA-binding protein 4).  
 GN WRKY4 OR AT1G13960 OR F7A19.5.  
 OS Arabidopsis thaliana (Mouse-ear cress).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;  
 OC eucosids 11; Brassicales; Brassicaceae; Arabidopsids.  
 OC NCBI\_TaxID=3702;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Chen C., Yu D., Du L., Fan B., Chen Z.;  
 RT "Identification of pathogen- and salicylic acid-induced WRKY  
 RT DNA-binding activities and genes encoding WRKY DNA-binding proteins in  
 RT Arabidopsis.";  
 RL Submitted (JAN-2000) to the EMBL/Genbank/DBJ databases.  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=cv. Columbia; TISSUE=flower;  
 RA Ulker B., Kuschni S., Somsich I.B.;  
 RT "Arabidopsis thaliana transcription factor WRKY4.";  
 RL Submitted (SEP-2001) to the EMBL/Genbank/DBJ databases.  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=cv. Columbia;  
 RX MEDLINE=21016719; PubMed=11130712;

RA Theologis A., Ecker J.R., Palm C.J., Federspiel N.A., Kaul S.,  
 RA White O., Alonso J., Altati H., Araujo R., Bowman C.L., Brooks S.Y.,  
 RA Buchler E., Chan A., Chao Q., Chen H., Cheuk R.F., Chin C.W.,  
 RA Chung M.K., Conn L., Conway A.B., Conway A.R., Creasy T.H., Dewar K.,  
 RA Dunn P., Elgu P., Feldlyum T.V., Feng J.-D., Fong B., Fujii C.Y.,  
 RA Gill J.E., Goldsmith A.D., Haas B., Hansen N.F., Hughes B., Huizar L.,  
 RA Hunter J.L., Jenkins J., Johnson-Hopson C., Khan S., Khaykin E.,  
 RA Kim C.J., Koo H.L., Kremenetskaia I., Kurtz D.B., Kwan A., Lam B.,  
 RA Langin-Hooper S., Lee A., Lee J.M., Lenz C.A., Li J.H., Li Y.-P.,  
 RA Lin X., Liu S.X., Liu Z.A., Luros J.S., Maiti R., Marziani A.,  
 RA Miltiescher J., Miranda M., Nguyen M., Niernan W.C., Osborne B.I.,  
 RA Pal G., Peterson J., Pham P.K., Rizzo M., Rooney T., Rowley D.,  
 RA Sakano H., Salzberg S.W., Schwartz J.R., Shinn P., Southwick A.M.,  
 RA Sun H., Tallon L.J., Tambunga G., Toriumi M.J., Town C.D.,  
 RA Utecherback T., Van Aken S., Vaysberg M., Vysotskaya V.S., Walker M.,  
 RA Wu D., Yu G., Fraser C.M., Venter J.C., Davis R.W.;  
 RA "Sequence and analysis of chromosome 1 of the plant Arabidopsis  
 RA thaliana.";  
 RT Nature 408:816-820(2000).  
 RL [4]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=cv. Columbia;  
 RN RN Shinzaki K., Davis R.W., Ecker J.R., Theologis A.;  
 RA "Riken Arabidopsis full length cDNA clones (RAPF) sequenced by the  
 RT SSP consortium (Salk/Stanford/PGSC).";  
 RL Submitted (JUL-2001) to the EMBL/Genbank/DBJ databases.  
 CC -1- SUBCELLULAR LOCATION: Nuclear (Potential).  
 CC -1- SIMILARITY: CONTAINS 2 WRKY DOMAINS.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.ebi.ac.uk/announcement/>  
 CC or send an email to [license@ebi.ac.uk](mailto:license@ebi.ac.uk)).

CC -----  
 DR EMBL; AF224703; AAK8313.1; -;  
 DR EMBL; AF425835; AAL13048.1; -;  
 DR EMBL; AC007576; AAD39282.1; -;  
 DR EMBL; AY045676; AAK74034.1; -;  
 DR InterPro; IPR003657; WRKY.  
 DR Pfam; PF03106; WRKY. 2.  
 DR PROSITE; PS50811; WRKY. 2.  
 KW Transcription regulation; DNA-binding; Nuclear protein; Repeat.  
 FT DNA\_BIND 223 287 WRKY 1.  
 FT DNA\_BIND 403 468 WRKY 2.  
 FT CONFLICT 112 147 AVALDLCNI (IN REF. 1 AND 3).  
 SQ SEQUENCE 514 AA; 55815 MM; 01010F8745C420C5 CRC64;

Query Match 42.2%; Score 38; DB 1; Length 514;  
 Best Local Similarity 36.4%; Pred. No. 92;  
 Matches 8; Conservative 5; Mismatches 3; Indels 6; Gaps 1;

QY 1 VGVG-----RDDNDRDPKRN 16  
 | | | | |  
 Db 354 VGNGETDVREKDENEPDKRS 375

RESULT 35  
 AHPE\_XANCH STANDARD; PRT; 530 AA.  
 ID AHPE\_XANCH  
 AC O06465;  
 DT 15-JUL-1999 (Rel. 38, Created)  
 DT 15-JUL-1999 (Rel. 38, Last sequence update)  
 DT 15-JUL-1999 (Rel. 38, Last annotation update)  
 DE Alky1 hydroperoxide reductase subunit F (EC 1.6.4.-).  
 GN AHPE.  
 OS Xanthomonas campestris (pv. phaseoli).  
 OC Bacteria; Proteobacteria; gamma subdivision; Xanthomonas group;  
 OC Xanthomonas.  
 OK NCBI\_TaxID=29445;

```

RN [1]
RX MEDLINE=9111968; PubMed=1989381;
RA Theilmann D.A., Stewart S.;
RT "Identification and characterization of the IE-1 gene of Orygia
RL pseudotsugata multicapsid nuclear polyhedrosis virus.";
RT Virology 180:492-508 (1991).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=97271300; PubMed=9126251;
RA Ahrens C.H., Russell R.R., Funk C.J., Evans J., Harwood S.,
RT Rohrmann G.F.;
RT "The sequence of the Orygia pseudotsugata multinnucleocapsid nuclear
RL polyhedrosis virus genome.";
RL Virology 229:381-399 (1997).
CC -1- FUNCTION: REGULATORY TRANSCRIPTIONAL PROTEIN, WHICH TRANS-
CC ACTIVATES GENE EXPRESSION FROM EARLY BACULOVIRUS PROMOTERS. CAN
CC ALSO TRANS-ACTIVATE ITS OWN PROMOTER, SUGGESTING THAT IT IS
CC AUTOREGULATED DURING NORMAL INFECTION OF INSECT CELLS.
CC -1- ALTERNATIVE PRODUCTS: A LARGER TRANSCRIPT MAY ARISE BY ALTERNATIVE
CC SPLICING OF THE SAME GENE.
CC -1- SIMILARITY: TO OTHER BACULOVIRUSES IE-1.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
DR EMBL; U94336; AAC45426.1; -.
DR HSSP; P09625; 1JDE.
DR InterPro; IPR000759; Adnrx_reductase.
DR InterPro; IPR001327; FAD_pyr_redux.
DR InterPro; IPR001100; Pyr_redux.
DR InterPro; IPR001003; Pyridine_redux_2.
DR InterPro; IPR003042; Rag_mnoxygenase.
DR Pfam; PF00070; Pyr_redux_1.
DR PRINTS; PR00419; ADKRDPAE.
DR PRINTS; PR00368; FADPNR.
DR PRINTS; PR00411; PNDRTASEI.
DR PRINTS; PR00469; PNDRTASEII.
DR PRINTS; PR00420; RNMNNOXGNASE.
DR Prodom; PD000139; FAD_pyr_redux_1.
DR PROSITE; PS00573; PYRIDINE_REDOX_2; 1.
DR Redox-active center; Oxidoreductase; NADP, NAD; Flavoprotein; FAD.
KW NP_BIND 214 229 PAD (ADP PART) (BY SIMILARITY).
FT DISULFID 344 347 REDOX-ACTIVE (BY SIMILARITY).
FT NP_BIND 356 370 NAD(P) (BY SIMILARITY).
FT NP_BIND 477 487 PAD (FLAVIN PART) (BY SIMILARITY).
SQ SEQUENCE 530 AA; 56921 MW; ABC1105C8A6FCD24 CRC64;

Query Match 42.2%; Score 38; DB 1; Length 530;
Best Local Similarity 77.8%; Pred. No. 96;
Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 5 RDDNDRDPSS 13
DB 56 RDDNDRDPSS 64

RESULT 36
TATR_NPVOV STANDARD; PRT; 560 AA.
ID TATR_NPVOV
AC P2214;
DT 01-AUG-1991 (Rel. 19, Created)
DT 01-AUG-1991 (Rel. 19, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE Trans-activating transcriptional regulatory protein (Immediate early
DE protein 1) (IE-1).
GN IE1.
OS Orygia pseudotsugata multicapsid polyhedrosis virus (OpMV).
OC Viruses; dsDNA viruses, no RNA stage; Baculoviridae;
OC Nucleopolyhedrovirus.
OX NCBI_TaxID=164623;
RN [1]
RP SEQUENCE FROM N.A.

```

```

RX MEDLINE=9111968; PubMed=1989381;
RA Theilmann D.A., Stewart S.;
RT "Identification and characterization of the IE-1 gene of Orygia
RL pseudotsugata multicapsid nuclear polyhedrosis virus.";
RT Virology 180:492-508 (1991).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=97271300; PubMed=9126251;
RA Ahrens C.H., Russell R.R., Funk C.J., Evans J., Harwood S.,
RT Rohrmann G.F.;
RT "The sequence of the Orygia pseudotsugata multinnucleocapsid nuclear
RL polyhedrosis virus genome.";
RL Virology 229:381-399 (1997).
CC -1- FUNCTION: REGULATORY TRANSCRIPTIONAL PROTEIN, WHICH TRANS-
CC ACTIVATES GENE EXPRESSION FROM EARLY BACULOVIRUS PROMOTERS. CAN
CC ALSO TRANS-ACTIVATE ITS OWN PROMOTER, SUGGESTING THAT IT IS
CC AUTOREGULATED DURING NORMAL INFECTION OF INSECT CELLS.
CC -1- ALTERNATIVE PRODUCTS: A LARGER TRANSCRIPT MAY ARISE BY ALTERNATIVE
CC SPLICING OF THE SAME GENE.
CC -1- SIMILARITY: TO OTHER BACULOVIRUSES IE-1.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
DR EMBL; M63414; AAA46700.1; -.
DR EMBL; U75930; AAC59144.1; -.
DR PIR; A38544; RGNVPM.
DR InterPro; IPR005092; TATR.
DR Pfam; PF03430; TATR_1.
KW Early protein; Transcription regulation; Trans-acting factor;
KW Alternative splicing.
FT DOMAIN 1 132 ASP/GLU-RICH (ACIDIC).
SQ SEQUENCE 560 AA; 64315 MW; 3991D4261A489BDD CRC64;

Query Match 42.2%; Score 38; DB 1; Length 560;
Best Local Similarity 53.8%; Pred. No. 1e+02;
Matches 7; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

QY 3 YGRDNDRDPSSK 15
DB 422 YGRDNDNDALAKK 434

RESULT 37
FMR1_XENLA STANDARD; PRT; 564 AA.
ID FMR1_XENLA
AC P51113;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Fragile X mental retardation protein 1 homolog (Protein FMR-1) (FMRP).
GN FMR1.
OS Xenopus laevis (African clawed frog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae;
OC Xenopodinae; Xenopus.
OX NCBI_TaxID=8355;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=95300772; PubMed=7781595;
RA Sloni M.C., Sloni H., Sauer W.H., Srinivasan S., Nussbaum R.L.,
RT Dreyfus G.;
RT "FXR1, an autosomal homolog of the fragile X mental retardation
RL gene";
RL EMBO J. 14:2401-2408 (1995).
CC -1- FUNCTION: RNA-BINDING PROTEIN. INTERACTS WITH FXR1 AND FXR2.
CC ASSOCIATED TO POLYSOMES AND MIGHT BE INVOLVED IN THE TRANSPORT OF
CC mRNA FROM THE NUCLEUS TO THE CYTOPLASM (BY SIMILARITY).

```

```
CC -1- SUBCELLULAR LOCATION: Cytoplasmic and nuclear (By similarity).
CC -1- SIMILARITY: BELONGS TO THE FMR1 FAMILY.
CC -1- SIMILARITY: CONTAINS 2 KH DOMAINS.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
DR EMBL; U25164; AAC59683.1; -.
DR HSSP; Q06787; 2FMR.
DR InterPro; IPR004087; KH dom.
DR InterPro; IPR004088; KH_type_1.
DR Pfam; PF00013; KH-domain; 2.
DR SMART; SM00322; KH; 2.
DR PROSITE; PS50084; KH_TYPE_1; 2.
KW Nuclear protein; RNA-binding; Repeat.
FT DOMAIN 222 251 KH 1.
FT DOMAIN 285 314 KH 2.
FT SITE 468 481 RNA-BINDING (RGC-BOX).
SQ SEQUENCE 564 AA; 63859 MW; 766379B83C24DE8F CRC64;

Query Match 42.2%; Score 38; DB 1; Length 564;
Best Local Similarity 47.4%; Pred. No. 1e+02;
Matches 9; Conservative 2; Mismatches 4; Indels 4; Gaps 1;

QY 2 GYGRDNDPRDPRK 16
Db 54 GYNKDINERDEVEVYSRAN 72

RESULT 38
ID DBP9_YEAST STANDARD; PRT; 594 AA.
AC Q06218;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Probable ATP-dependent RNA helicase DBP9 (DEAD-box protein 9).
GN DBP9 OR YLR276C OR L9328.3.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomyces.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=S288C / AB972;
RA Johnston M., Andrews S., Brinkman R., Cooper J., Ding H., Du Z.,
RA Favella A., Fulton L., Gattung S., Greco T., Kirsten J.,
RA Kucaba T., Hallsworth K., Hawkins J., Hillier L., Jier M.,
RA Johnson D., Johnston L., Langston Y., Latreille P., Le T.,
RA Madris E., Meneses S., Miller N., Nham N., Pauley A., Peluso D.,
RA Rikken L., Riles L., Taich A., Trevaaskis E., Vignati D.,
RA Wilcox L., Wohlman P., Vaudin M., Wilson R., Waterson R.;
RA Submitted (NOV-1994) to the EMBL/Genbank/DBJ databases.
RL [2]
RP CHARACTERIZATION.
RX MEDLINE=21449431; PubMed=11565753;
RA Dagenorn M.-C., Kressler D., Linder P.;
RT DBP9p, a putative ATP-dependent RNA helicase involved in
RT 60S-ribosomal-subunit biogenesis, functionally interacts with dbp9p.;
RL RNA 7.1317-1334(2001).
CC -1- FUNCTION: Probable ATP-binding RNA helicase involved in the
CC biogenesis of 60S ribosomal subunits and is required for the
CC normal formation of 25S and 5.8S rRNAs.
CC -1- SUBUNIT: Interacts with DBP6.
CC -1- SUBCELLULAR LOCATION: Nuclear; nucleolar.
CC -1- SIMILARITY: BELONGS TO THE DEAD BOX HELICASE FAMILY.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
```

```
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
DR EMBL; U17245; AAB67366.1; -.
DR HSSP; Q58083; 1HV8.
DR SGD; S0004266; DBP9.
DR InterPro; IPR001410; DEAD.
DR InterPro; IPR000629; DEAD_box.
DR InterPro; IPR001650; Helicase_C.
DR Pfam; PF00270; DEAD; 1.
DR Pfam; PF00271; helicase_C; 1.
DR SMART; SM00487; DEXDC; 1.
DR SMART; SM00480; HELIC_C; 1.
DR PROSITE; PS00039; DEAD_ATP_HELICASE, FALSE_NEG.
KW Hydrolyase; Helicase; ATP-binding; RNA-binding; Nuclear protein;
KW rRNA processing; Ribosome biogenesis.
FT NP BIND 62 69 ATP (POTENTIAL).
FT SITE 179 182 DEAD BOX.
SQ SEQUENCE 594 AA; 68059 MW; CCE8E0AF9B4071D7 CRC64;

Query Match 42.2%; Score 38; DB 1; Length 594;
Best Local Similarity 50.0%; Pred. No. 1.1e+02;
Matches 6; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

QY 4 GRDNDPRDPRSRK 15
Db 359 GEPENDKKPSRK 370

RESULT 39
ID PRIM_LISIN STANDARD; PRT; 626 AA.
AC Q92B05;
DT 15-JUN-2002 (Rel. 41, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE DNA primase (EC 2.7.7.-).
GN DNAG OR LIN1492.
OS Listeria innocua.
OC Bacteria; Firmicutes; Bacillales; Listeriaceae; Listeria.
OX NCBI_TaxID=1642;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CLIP 11262 / Serovar 6a;
RX MEDLINE=21537279; PubMed=11679669;
RA Glaser P., Frangeul L., Buchrieser C., Rusniok C., Amend A.,
RA Baquero F., Berche P., Bloeker H., Brandt P., Chakraborty T.,
RA Charbit A., Cherouni F., Couve E., de Darvar A., Dehnot P.,
RA Domani E., Dominguez-Bernal G., Duchaud E., Durant L., Dussurget O.,
RA Entian K.-D., Fshiz H., Garcia-del Portillo F., Garrido P.,
RA Gautier L., Goebel W., Gomez-Lopez N., Hahn T., Haul J., Jackson D.,
RA Jones L.-M., Kaerst U., Kreft J., Kuhn M., Kunst F., Kuxupak G.,
RA Madueno E., Maizouram A., Mata Vicente J., Ng E., Nedjari H.,
RA Nardusiek G., Novella S., de Pablos B., Perez-Diaz J.-C., Puccell R.,
RA Remmel B., Rose M., Schueter T., Simoes N., Tierrez A.,
RA Vazquez-Boland J.-A., Voss H., Weiland J., Cossart P.;
RL Science 294:849-852(2001).
CC -1- FUNCTION: DNA PRIMASE IS THE POLYMERASE THAT SYNTHESIZES SMALL
CC REPLICATION FORKS DURING CHROMOSOMAL DNA SYNTHESIS.
CC -1- SUBUNIT: BINDS ONE ZINC ION PER MOLECULE (BY SIMILARITY).
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
```



CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

DR EMBL; AL596168; CAC96723.1; -  
 DR Listlist; L1N01492; -  
 DR InterPro; IPR002936; DNAPrim\_toprim.  
 DR InterPro; IPR002694; Znf\_CHC2.  
 DR Pfam; PF01751; Toprim; 1.  
 DR Pfam; PF01807; zf-CHC2; 1.  
 DR ProDom; PD002988; Znf-CHC2; 1.  
 KM Transferrase; DNA replication; DNA-directed RNA polymerase; Primosome;  
 FT Zinc-finger; Zinc; Metal-binding; Complete proteome.  
 FT ZN\_FING 39 63 CHC2-TYPE (BY SIMILARITY).  
 SQ SEQUENCE 626 AA; 71801 MM; 1E708649628F2346 CRC64;

Query March 42.2%; Score 38; DB 1; Length 626;  
 Best Local Similarity 42.9%; Pred. No. 1.2e+02;  
 Matches 6; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

QY 1 VGYGRDNDRDPSR 14  
 :|||:|||||:  
 Db 527 IGYFAEGNDADPHK 540

RESULT 40  
 YH5 YEAST STANDARD; PRT; 659 AA.  
 ID YH5 YEAST  
 AC P38760;  
 DT 01-FEB-1995 (Rel. 31, Created)  
 DT 01-FEB-1995 (Rel. 31, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Hypothetical 75.9 kDa protein in SPO13-ARG4 intergenic region.  
 GN YHR015W.  
 OS Saccharomyces cerevisiae (Baker's yeast).  
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
 OC Saccharomycetales; Saccharomycetaceae; Saccharomyces.  
 OX NCBI\_TaxID=4932;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=5288C / AB972;  
 RX MEDLINE=94378003; PubMed=8091229;  
 RA Johnston M., Andrews S., Brinkman R., Cooper J., Ding H., Dover J.,  
 RA Du Z., Favell A., Fulton L., Gattung S., Geisel C., Kilsten J.,  
 RA Kucaba T., Hillier L., Jier M., Johnston L., Langston Y.,  
 RA Latreille P., Louis E.J., Macri C., Mardis E., Menezes S., Mouser L.,  
 RA Nhan M., Rifkin L., Riles L., St Peter H., Trevaatis E., Vaughan K.,  
 RA Vignati D., Wilcox L., Wohlman P., Waterston K., Wilson R.,  
 RA Vaudin M.;  
 RT "Complete nucleotide sequence of Saccharomyces cerevisiae chromosome  
 VIII.";  
 RL Science 265:2077-2082(1994).  
 CC -|- SIMILARITY: CONTAINS 3 RNA RECOGNITION MOTIFS (RRM).  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

DR EMBL; U10400; AAB68942.1; -  
 DR PIR; S46788; S46788.  
 DR HSSP; P11940; ICVJ.  
 DR SGD; S0001057; YHR015W.  
 DR InterPro; IPR000504; RNA\_rec\_mot.  
 DR InterPro; IPR003955; RRM\_2.  
 DR Pfam; PF00076; rrm; 3.  
 DR Pfam; PF00360; RRM; 3.  
 DR SMART; SM00362; RRM\_2; 1.  
 DR SMART; SM00362; RRM; 3.  
 DR PROSITE; PS0102; RRM; 3.  
 DR PROSITE; PS00030; RRM\_RNP\_1; FALSE NEG.  
 KM Hypothetical protein; RNA-binding; Repeat.

FT DOMAIN 111 189 RNA-BINDING (RRM) 1.  
 FT DOMAIN 199 267 RNA-BINDING (RRM) 2.  
 FT DOMAIN 313 389 RNA-BINDING (RRM) 3.  
 SQ SEQUENCE 659 AA; 75919 MM; DF8E3F4D12A63871 CRC64;

Query March 42.2%; Score 38; DB 1; Length 659;  
 Best Local Similarity 42.1%; Pred. No. 1.2e+02;  
 Matches 8; Conservative 4; Mismatches 3; Indels 4; Gaps 1;

QY 1 VGYG----RDNDRDPSRK 15  
 :|||:|||||:  
 Db 152 LGYGYNFKDNDADPSARK 170

Search completed: April 9, 2003, 12:56:06  
 Job time : 6.80328 secs

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.4 p5.4578  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 9, 2003, 12:50:42 ; Search time 14.1639 Seconds  
(without alignments)  
232.757 Million cell updates/sec

Title: US-10-020-441-3

Perfect score: 90  
Sequence: 1 VGYGRDNDNRDPSRKN 16

Scoring table: BLOSUM62  
Gapop 10.0, Gapext 0.5

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%  
Listing first 45 summaries

Database :

1: SPREMBL\_21:\*  
2: sp\_archaea:\*  
3: sp\_bacteria:\*  
4: sp\_fungi:\*  
5: sp\_human:\*  
6: sp\_invertebrate:\*  
7: sp\_mammal:\*  
8: sp\_mhc:\*  
9: sp\_organelle:\*  
10: sp\_phage:\*  
11: sp\_plant:\*  
12: sp\_rodent:\*  
13: sp\_virus:\*  
14: sp\_vertebrate:\*  
15: sp\_unclassified:\*  
16: sp\_rvirus:\*  
17: sp\_bacteriap:\*  
17: sp\_archaeap:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	90	100.0	274	5	Q26552 echistosoma
2	81	90.0	257	5	Q26553 schistosoma
3	51	56.7	385	16	Q985A4 rhizobium 1
4	47	52.2	144	13	Q92167 xenopus lae
5	47	52.2	363	12	Q9J8C2 spodopora
6	47	52.2	1826	4	Q9N0R8
7	46	51.1	407	10	Q64907 pelargonium
8	46	51.1	1428	3	Q8TG58 saccharomyc
9	46	51.1	1428	3	Q8TG56 saccharomyc
10	46	51.1	1428	3	Q8TG53 saccharomyc
11	46	51.1	1428	3	Q8TG47 saccharomyc
12	46	51.1	1428	3	Q8TG46 saccharomyc
13	46	51.1	1428	3	Q8TG44 saccharomyc
14	46	51.1	1428	3	Q8TG43 saccharomyc
15	46	51.1	1428	3	Q8TF86 saccharomyc
16	45.5	50.6	388	5	Q9GZFO caenorhabdi

17	45	50.0	209	10	Q9FG28 arabidopsis
18	45	50.0	593	2	Q93QU8 escherichia
19	45	50.0	716	10	Q9CSH5 arabidopsis
20	45	50.0	716	10	Q93ZH4 arabidopsis
21	45	50.0	810	10	Q9ZWI3 cucurbita m
22	44	48.9	165	5	Q95QI9 caenorhabdi
23	44	48.9	394	2	Q9R6A3 agrobacteri
24	44	48.9	443	10	Q9LMM5 arabidopsis
25	44	48.9	515	5	Q8TIN3 dictyosteli
26	44	48.9	561	5	Q09542 caenorhabdi
27	44	48.9	8243	5	Q96554 cryptospori
28	43	47.8	139	10	Q94HP4 oryza activ
29	43	47.8	142	13	Q919J2 xenopus lae
30	43	47.8	143	13	Q919J3 xenopus lae
31	43	47.8	219	10	Q9FIE4 arabidopsis
32	43	47.8	401	13	Q9PRG7 xenopus lae
33	43	47.8	414	13	Q90Z89 xenopus lae
34	43	47.8	419	2	Q06860 shewanella
35	43	47.8	444	5	Q9VD47 agrobacteri
36	43	47.8	477	4	Q969S3 homo sapien
37	43	47.8	523	2	Q9WVT7 agrobacteri
38	43	47.8	536	16	Q8UEA8 agrobacteri
39	43	47.8	570	10	Q24041 litidendro
40	43	47.8	686	16	Q9AEY0 caulobacter
41	42	46.7	150	16	Q8VQ47 mycobacteri
42	42	46.7	174	12	Q91F42 cydia pomon
43	42	46.7	215	5	Q9VSR9 dirosophila
44	42	46.7	223	5	Q8SVY2 dirosophila
45	42	46.7	461	10	Q9FG27 arabidopsis

## ALIGNMENTS

### RESULT 1

ID	Q26552	PRELIMINARY;	PRT;	274 AA.
AC	Q26552;			
DT	01-NOV-1996 (TREMURel. 01, Created)			
DT	01-NOV-1996 (TREMURel. 01, Last sequence update)			
DT	01-MAR-2002 (TREMURel. 20, Last annotation update)			
DE	Elastase.			
OS	Schistosoma mansoni (Blood fluke).			
OC	Eukaryota; Metazoa; Platyhelminthes; Trematoda; Digenea; Strigeididae;			
OC	Schistosomatidae; Schistosomatidae; Schistosoma.			
OX	NCBI_Taxid=6183;			
RX	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=PUERTO RICO.			
RX	MEDLINE=96362066; PubMed=8720180;			
RA	Pierrot C., Capron A., Khalife J.;			
RT	"Cloning and characterization of two genes encoding Schistosoma			
RT	mansoni elastase.";			
RL	Mol. Biochem. Parasitol. 75:113-117(1995).			
CC	-1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE			
CC	TRYPSIN FAMILY.			
DR	EMBL; U31768; AAC46967.1; -.			
DR	HSSP; P20231; IIAO.			
DR	MEROPS; S01.144; -.			
DR	InterPro; IPR001314; Chymotrypsin.			
DR	InterPro; IPR001254; Ser protease_Try.			
DR	Pfam; PF00089; trypsin_1.			
DR	PRINTS; PR00722; CHYMOTRYPSIN.			
DR	SMART; SM00020; TRYP_SPC; 1.			
DR	PROSITE; PS50240; TRYPsin_DOM; 1.			
DR	PROSITE; PS00134; TRYPsin_HIS; UNKNOWN_1.			
KW	Hydrolase; Serine protease.			
SO	SEQUENCE 274 AA; 29586 MW; 53F21914FDDDA450 CRC64;			

Query Match 100.0%; Score 90; DB 5; Length 274;  
Best Local Similarity 100.0%; Pred. No. 4.3e-07;  
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 VGYGRDNDNDPDRSK 16  
Db 168 VGYGRDNDNDPDRSK 183

## RESULT 2

ID Q26553 PRELIMINARY; PRT; 257 AA.  
AC Q26553;  
DT 01-NOV-1996 (TREMBLrel. 01, Created)  
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)  
DT 01-MAR-2002 (TREMBLrel. 20, Last annotation update)  
DE Elastase.  
OS Schistosoma mansoni (Blood fluke).  
OC Eukaryota; Metazoa; Platyhelminthes; Trematoda; Digenea; Strigeidida;  
OC Schistosomatidae; Schistosomatidae; Schistosoma.  
OX NCBI\_TaxID=6183;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=PUERTO RICO;  
RX MEDLINE=96362066; PubMed=8720180;  
RA Pierrot C., Capron A., Khalife J.;  
RT "Cloning and characterization of two genes encoding Schistosoma  
manson elastase.";  
RL Mol. Biochem. Parasitol. 75:113-117(1995).  
DR HSSP; P00766; 1GCD.  
DR EMBL; U31769; AAC46968.1; -.  
DR InterPro; IPR001254; Set.protease\_Try.  
DR Pfam; PF00089; trypsin\_1.  
DR SMART; SM00020; Tryp\_Spe; 1.  
DR PROSITE; PS50240; TRYPsin\_DOM; 1.  
DR PROSITE; PS00134; TRYPsin\_His; UNKNOWN\_1.  
KW Hydrolase; Serine protease.  
SQ SEQUENCE 257 AA; 27971 MW; 00F8872799F9D64 CRC64;

Query Match 90.0%; Score 81; DB 5; Length 257;  
Best Local Similarity 93.3%; Pred. No. 1.3e-05;  
Matches 14; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 VGYGRDNDNDPDRSK 15  
Db 164 VGYGRDNDNDPDRSK 178

## RESULT 3

ID Q985A4 PRELIMINARY; PRT; 365 AA.  
AC Q985A4;  
DT 01-OCT-2001 (TREMBLrel. 18, Created)  
DT 01-OCT-2001 (TREMBLrel. 18, Last sequence update)  
DT 01-MAR-2002 (TREMBLrel. 20, Last annotation update)  
DE Porin.  
GN MUR7766.  
OS Rhizobium loti (Mesorhizobium loti).  
OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;  
OC Phyllobacteriaceae; Mesorhizobium.  
OX NCBI\_TaxID=381;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=MAFF303099;  
RX MEDLINE=21082930; PubMed=11214968;  
RA Kaneke T., Nakamura Y., Sato S., Asamizu E., Kato T., Sasamoto S.,  
RA Watanabe A., Idegawa K., Ishikawa A., Kawashima K., Kimura T.,  
RA Kohida Y., Kiyokawa C., Kohara M., Matsumoto M., Matsuno A.,  
RA Mochizuki Y., Nakayama S., Nakazaki N., Shimo S., Sugimoto M.,  
RA Takeuchi C., Yamada M., Tabata S.;  
RT "Complete genome structure of the nitrogen-fixing symbiotic bacterium  
Mesorhizobium loti.";  
RL DNA Res. 7:331-338(2000).  
EMBL; AP003012; BAB54159.1; -.  
DR InterPro; IPR003684; Porin\_alphaabac.  
DR Pfam; PF02530; Porin\_2; 1.  
KW Complete proteome.

SQ SEQUENCE 365 AA; 41848 MW; 83838C0D9DD615E1 CRC64;  
Query Match 56.7%; Score 51; DB 16; Length 385;  
Best Local Similarity 75.0%; Pred. No. 2.2;  
Matches 9; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Qy 2 GYGRDNDNDPDRSK 13  
Db 272 GYGRDNDNDPDRSK 283

## RESULT 4

ID Q92167 PRELIMINARY; PRT; 144 AA.  
AC Q92167;  
DT 01-NOV-1996 (TREMBLrel. 01, Created)  
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)  
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)  
DE FROG oviduct specific protein-1B (Fragment).  
GN FOSP-1B.  
OS Xenopus laevis (African clawed frog).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipidea; Pipidae;  
OC Xenopodinae; Xenopus.  
OX NCBI\_TaxID=8355;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=OVIDUCT;  
RX MEDLINE=9526271; PubMed=7744034;  
RA Mehta R.J., Tata J.R.;  
RT "Structural and functional characterization and cloning of Xenopus  
RT FOSP-1 (frog oviduct-specific protein-1) gene.";  
RL Eur. J. Biochem. 229:224-232(1995).  
DR EMBL; X81570; CAA57266.1; -.  
DR EMBL; X81572; CAA57268.1; -.  
DR PRINTS; PR01574; TUBBYPROTEIN.  
FT NON TER 144 144  
SQ SEQUENCE 144 AA; 16018 MW; 401AAD8C7D2B50FC CRC64;

Query Match 52.3%; Score 47; DB 13; Length 144;  
Best Local Similarity 53.3%; Pred. No. 3.4;  
Matches 8; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

Qy 2 GYGRDNDNDPDRSK 16  
Db 91 GYGRDNDNDPDRSK 105

## RESULT 5

ID Q9J8C2 PRELIMINARY; PRT; 363 AA.  
AC Q9J8C2;  
DT 01-OCT-2000 (TREMBLrel. 15, Created)  
DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)  
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)  
DE ORF13 38..7kd.  
OS Spodoptera exigua nucleopolyhedrovirus.  
OC Viruses; dsDNA viruses, no RNA stage; Baculoviridae;  
OC Nucleopolyhedrovirus.  
OX NCBI\_TaxID=10454;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=20036646; PubMed=10567663;  
RA IJkel W.F., van Strien E.A., Helens J.G., Broer R., Zuidema D.,  
RA Goldbach R.W., Vlak J.M.;  
RT "Sequence and organization of the spodoptera exigua multicapsid  
RT nucleopolyhedrovirus genome.";  
RL J. Gen. Virol. 80:3289-3304(1999).  
RN [2]  
RP SEQUENCE FROM N.A.  
RA IJkel W.F.J., van Strien E.A., Helens J.G.M., Broer R., Zuidema D.,  
RA Goldbach R.W., Vlak J.M.;  
RT Submitted (JUL-1999) to the EMBL/GenBank/DBJ databases.

DR EMBL; AF169823; AAF3543.1; -.  
 DR InterPro; IPR003497; BRO\_Nterm.  
 DR Pfam; PF02498; BRO; 1.  
 SQ SEQUENCE 363 AA; 42338 MW; EDD6AE6EBE02DFD9 CRC64;

Query Match 52.2%; Score 47; DB 12; Length 363;  
 Best Local Similarity 60.0%; Pred. No. 9.8;  
 Matches 9; Conservative 1; Mismatches 5; Indels 0; Gaps 0;

OY 2 GYGRDNDPDRSKN 16  
 DB 11 GYFDDDDSDERKN 25

## RESULT 6

O9NQOT8 PRELIMINARY; PRT; 1826 AA.

AC O9NQOT8  
 DT 01-OCT-2000 (TREMBlrel. 15; Created)  
 DT 01-OCT-2000 (TREMBlrel. 15; Last sequence update)  
 DT 01-JUN-2002 (TREMBlrel. 21; Last annotation update)  
 DE Kinesin-like protein GAKIN.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 CX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=20435838; PubMed=10859302;  
 RA Hanada T., Lin L., Tibaldi E.V., Reinherz E.L., Chishti A.H.,  
 RT "GAKIN: a novel kinesin-like protein associates with the human  
 RT homologue of the Drosophila discs large tumor suppressor in T  
 RT lymphocytes.";   
 RL J. Biol. Chem. 275:28774-28784(2000).  
 DR EMBL; AF279865; AAF81263.1; -.  
 DR HSSP; P17119; 3KAR.  
 DR InterPro; IPR000938; CAP\_Gly.  
 DR InterPro; IPR000253; FHA domain.  
 DR InterPro; IPR001752; kinesin\_motor.  
 DR Pfam; PF01302; CAP\_GLY; 1.  
 DR Pfam; PF00498; FHA; 1.  
 DR Pfam; PF00225; kinesin; 1.  
 DR PRINTS; PR00380; KINESINHEAVY.  
 DR SMART; SMO0240; FHA; 1.  
 DR SMART; SMO0129; KISC; 1.  
 DR PROSITE; PS00845; CAP\_GLY\_1; 1.  
 DR PROSITE; PS00411; KINESIN\_MOTOR\_DOMAIN1; 1.  
 DR PROSITE; PS50067; KINESIN\_MOTOR\_DOMAIN2; 1.  
 DR ATP-binding; Coiled coil; Microtubules; Motor protein.  
 SQ SEQUENCE 1826 AA; 202665 MW; C614E73A89E89ED CRC64;

Query Match 52.2%; Score 47; DB 4; Length 1826;  
 Best Local Similarity 66.7%; Pred. No. 61;  
 Matches 8; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

OY 5 RDDNDPDRSKN 16  
 DB 553 REDDDPDRSKN 564

## RESULT 7

O64907 PRELIMINARY; PRT; 407 AA.

AC O64907  
 DT 01-AUG-1998 (TREMBlrel. 07; Created)  
 DT 01-AUG-1998 (TREMBlrel. 07; Last sequence update)  
 DT 01-JUN-2002 (TREMBlrel. 21; Last annotation update)  
 DE Omega-3 desaturase (Fragment).  
 GN PKH-15.  
 OS Pelargonium hortorum.  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;  
 OC Geraniales; Geraniaceae; Pelargonium.

OX NCBI\_TaxID=4031;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Schultz D.J., Mamma R.O., Cox-Foster D., Craig R., Medford J.I.;  
 RT "Geranium omega-3 desaturase.";   
 RL Submitted (Aug-1997) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AF020204; AAC16443.1; -.  
 DR InterPro; IPR001225; FA\_desaturase.  
 DR Pfam; PF00487; FA\_desaturase; 1.  
 DR ProDom; PD001081; FA\_desaturase; 2.  
 FT NON TER 1

SQ SEQUENCE 407 AA; 47169 MW; 94FC4C3DF02B7C4E CRC64;

Query Match 51.1%; Score 46; DB 10; Length 407;  
 Best Local Similarity 72.7%; Pred. No. 16;  
 Matches 8; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

OY 3 YGRDNDPDRPS 13  
 DB 41 YGHDDSDPDRPS 51

## RESULT 8

O8TG58 PRELIMINARY; PRT; 1428 AA.

AC O8TG58  
 DT 01-JUN-2002 (TREMBlrel. 21; Created)  
 DT 01-JUN-2002 (TREMBlrel. 21; Last sequence update)  
 DT 01-JUN-2002 (TREMBlrel. 21; Last annotation update)  
 DE TOP2.  
 GN Saccharomyces cerevisiae (Baker's yeast).  
 OS Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
 OC Saccharomycetales; Saccharomycetaceae; Saccharomyces.  
 CX NCBI\_TaxID=4932;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=YJM280;  
 RX MEDLINE=21904897; PubMed=11907579;  
 RA Steinmetz L.M., Sinha H., Richards D.R., Spiegelman J.I., Oefner P.J.,  
 RT McCusker J.H., Davis R.W.;  
 RT "Dissecting the architecture of a quantitative trait locus in yeast.";   
 RL Nature 416:326-330(2002).  
 DR EMBL; AF458970; AAM00524.1; -.  
 SQ SEQUENCE 1428 AA; 164285 MW; 9B3CA9671D1DD33 CRC64;

Query Match 51.1%; Score 46; DB 3; Length 1428;  
 Best Local Similarity 72.7%; Pred. No. 68;  
 Matches 8; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

OY 6 DDNDPDRSKN 16  
 DB 1198 DDEYDPSKRN 1208

## RESULT 9

O8TG56 PRELIMINARY; PRT; 1428 AA.

AC O8TG56  
 DT 01-JUN-2002 (TREMBlrel. 21; Created)  
 DT 01-JUN-2002 (TREMBlrel. 21; Last sequence update)  
 DT 01-JUN-2002 (TREMBlrel. 21; Last annotation update)  
 DE TOP2.  
 GN YNL088W.  
 OS Saccharomyces cerevisiae (Baker's yeast).  
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
 OC Saccharomycetales; Saccharomycetaceae; Saccharomyces.  
 CX NCBI\_TaxID=4932;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=YJM339;  
 RX MEDLINE=21904897; PubMed=11907579;  
 RA Steinmetz L.M., Sinha H., Richards D.R., Spiegelman J.I., Oefner P.J.,

RA McCusker J.H., Davis R.W.;  
RT "Dissecting the architecture of a quantitative trait locus in yeast."  
RL Nature 416:326-330(2002).  
DR EMBL; AF458973; AAM00542.1; "-  
SQ SEQUENCE 1428 AA; 164287 MW; 8E0754C98B61B6B CRC64;

Query Match 51.1%; Score 46; DB 3; Length 1428;  
Best Local Similarity 72.7%; Pred. No. 68;  
Matches 8; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

OY 6 DDNDRPSRKN 16  
Db 1198 DDEDYDPSKKN 1208

## RESULT 10

O8TG53 PRELIMINARY; PRT; 1428 AA.

AC O8TG53; 01-JUN-2002 (TRENBLrel. 21, Created)  
DT 01-JUN-2002 (TRENBLrel. 21, Last sequence update)  
DE 01-JUN-2002 (TRENBLrel. 21, Last annotation update)  
DE TOP2.

GN YNL088W.  
OS Saccharomyces cerevisiae (Baker's yeast).  
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
OC Saccharomycetales; Saccharomycetaceae; Saccharomyces.  
OX NCBI\_TaxID=4932;

RN [1]  
RP SEQUENCE FROM N.A.

RC STRAIN=YJM421;  
RX MEDLINE=21904897; PubMed=11907579;  
RA Steinmetz L.M., Sinha H., Richards D.R., Spiegelman J.I., Oefner P.J.,  
RA McCusker J.H., Davis R.W.;  
RT "Dissecting the architecture of a quantitative trait locus in yeast."  
RL Nature 416:326-330(2002).  
DR EMBL; AF458974; AAM00548.1; "-  
SQ SEQUENCE 1428 AA; 164273 MW; A8F6F7B99FEBC1 CRC64;

Query Match 51.1%; Score 46; DB 3; Length 1428;  
Best Local Similarity 72.7%; Pred. No. 68;  
Matches 8; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

OY 6 DDNDRPSRKN 16  
Db 1198 DDEDYDPSKKN 1208

## RESULT 11

O8TG47 PRELIMINARY; PRT; 1428 AA.

AC O8TG47; 01-JUN-2002 (TRENBLrel. 21, Created)  
DT 01-JUN-2002 (TRENBLrel. 21, Last sequence update)  
DE 01-JUN-2002 (TRENBLrel. 21, Last annotation update)  
DE TOP2.

GN YNL088W.  
OS Saccharomyces cerevisiae (Baker's yeast).  
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
OC Saccharomycetales; Saccharomycetaceae; Saccharomyces.  
OX NCBI\_TaxID=4932;

RN [1]  
RP SEQUENCE FROM N.A.

RC STRAIN=YJM1129;  
RX MEDLINE=21904897; PubMed=11907579;  
RA Steinmetz L.M., Sinha H., Richards D.R., Spiegelman J.I., Oefner P.J.,  
RA McCusker J.H., Davis R.W.;  
RT "Dissecting the architecture of a quantitative trait locus in yeast."  
RL Nature 416:326-330(2002).  
DR EMBL; AF458978; AAM00572.1; "-  
SQ SEQUENCE 1428 AA; 164200 MW; D2658F4DC9281BA7 CRC64;

Query Match 51.1%; Score 46; DB 3; Length 1428;

Best Local Similarity 72.7%; Pred. No. 68;  
Matches 8; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

OY 6 DDNDRPSRKN 16  
Db 1198 DDEDYDPSKKN 1208

## RESULT 12

O8TG46 PRELIMINARY; PRT; 1428 AA.

AC O8TG46; 01-JUN-2002 (TRENBLrel. 21, Created)  
DT 01-JUN-2002 (TRENBLrel. 21, Last sequence update)  
DE 01-JUN-2002 (TRENBLrel. 21, Last annotation update)  
DE TOP2.

GN YNL088W.  
OS Saccharomyces cerevisiae (Baker's yeast).  
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
OC Saccharomycetales; Saccharomycetaceae; Saccharomyces.  
OX NCBI\_TaxID=4932;

RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=YJM269;  
RX MEDLINE=21904897; PubMed=11907579;  
RA Steinmetz L.M., Sinha H., Richards D.R., Spiegelman J.I., Oefner P.J.,  
RA McCusker J.H., Davis R.W.;  
RT "Dissecting the architecture of a quantitative trait locus in yeast."  
RL Nature 416:326-330(2002).  
DR EMBL; AF458979; AAM00578.1; "-  
SQ SEQUENCE 1428 AA; 164250 MW; A89E8379793FE7A2 CRC64;

Query Match 51.1%; Score 46; DB 3; Length 1428;  
Best Local Similarity 72.7%; Pred. No. 68;  
Matches 8; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

OY 6 DDNDRPSRKN 16  
Db 1198 DDEDYDPSKKN 1208

## RESULT 13

O8TG44 PRELIMINARY; PRT; 1428 AA.

AC O8TG44; 01-JUN-2002 (TRENBLrel. 21, Created)  
DT 01-JUN-2002 (TRENBLrel. 21, Last sequence update)  
DE 01-JUN-2002 (TRENBLrel. 21, Last annotation update)  
DE TOP2.

GN YNL088W.  
OS Saccharomyces cerevisiae (Baker's yeast).  
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
OC Saccharomycetales; Saccharomycetaceae; Saccharomyces.  
OX NCBI\_TaxID=4932;

RN [1]  
RP SEQUENCE FROM N.A.

RC STRAIN=YJM270;  
RX MEDLINE=21904897; PubMed=11907579;  
RA Steinmetz L.M., Sinha H., Richards D.R., Spiegelman J.I., Oefner P.J.,  
RA McCusker J.H., Davis R.W.;  
RT "Dissecting the architecture of a quantitative trait locus in yeast."  
RL Nature 416:326-330(2002).  
DR EMBL; AF458980; AAM00584.1; "-  
SQ SEQUENCE 1428 AA; 164240 MW; 5E80D2AA9B984FB CRC64;

Query Match 51.1%; Score 46; DB 3; Length 1428;  
Best Local Similarity 72.7%; Pred. No. 68;  
Matches 8; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

OY 6 DDNDRPSRKN 16  
Db 1198 DDEDYDPSKKN 1208

```

RESULT 14
Q8TG43 PRELIMINARY; PRT; 1428 AA.
AC 08TG43;
DT 01-JUN-2002 (TREMBlrel. 21, Created)
DT 01-JUN-2002 (TREMBlrel. 21, Last sequence update)
DE 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
DE TOP2.
GN YNL088W.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomyces.
OX NCBI_TaxId=4932;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=YJM627;
RX MEDLINE=21904897; PubMed=11907579;
RA Steimetz L.M., Sinha H., Richards D.R., Spiegelman J.I., Oefner P.J.,
RA McCusker J.H., Davis R.W.;
RT "Dissecting the architecture of a quantitative trait locus in yeast.";
RL Nature 416:326-330(2002).
DR EMBL; AF458981; AM00590.1; -
SQ SEQUENCE 1428 AA; 164313 MW; 0E29FB9A88BA0286 CRC64;

Query Match 51.1%; Score 46; DB 3; Length 1428;
Best Local Similarity 72.7%; Pred. No. 68;
Matches 8; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 6 DDNDRDPSRKN 16
Db 1198 DDEYDPSKKN 1208

RESULT 15
Q8TF86 PRELIMINARY; PRT; 1428 AA.
AC 08TF86;
DT 01-JUN-2002 (TREMBlrel. 21, Created)
DT 01-JUN-2002 (TREMBlrel. 21, Last sequence update)
DE 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
DE TOP2.
GN YNL088W.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomyces.
OX NCBI_TaxId=4932;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=YJM789; SKI, AND W303;
RX MEDLINE=21904897; PubMed=11907579;
RA Steimetz L.M., Sinha H., Richards D.R., Spiegelman J.I., Oefner P.J.,
RA McCusker J.H., Davis R.W.;
RT "Dissecting the architecture of a quantitative trait locus in yeast.";
RL Nature 416:326-330(2002).
DR EMBL; AF458975; AM00554.1; -
DR EMBL; AF458976; AM00550.1; -
DR EMBL; AF458977; AM00566.1; -
SQ SEQUENCE 1428 AA; 164244 MW; E9AF2E3EDCEEF06 CRC64;

Query Match 51.1%; Score 46; DB 3; Length 1428;
Best Local Similarity 72.7%; Pred. No. 68;
Matches 8; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 6 DDNDRDPSRKN 16
Db 1198 DDEYDPSKKN 1208

RESULT 16
Q9GZFO PRELIMINARY; PRT; 388 AA.
AC Q9GZFO;

```

```

DT 01-MAR-2001 (TREMBlrel. 16, Created)
DT 01-MAR-2001 (TREMBlrel. 16, Last sequence update)
DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
DE Hypothetical 43.4 kDa protein.
GN F22F7.2.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditicoidea;
OC Rhabditidae; Peleoderinae; Caenorhabditis.
OX NCBI_TaxId=6239;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BRISTOL N2;
RX MEDLINE=99069613; PubMed=9851916;
RA None;
RT "Genome sequence of the nematode C. elegans: a platform for
investigating biology." The C. elegans Sequencing Consortium.
RL Science 282:2012-2018(1998).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=BRISTOL N2;
RA Miller N., Kramer J., Smith A.;
RT "The sequence of C. elegans cosmid F22F7."
RL Submitted (JUN-1998) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=BRISTOL N2;
RA Waterston R.;
RT "Direct Submission."
RL Submitted (SEP-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF067937; AAF99917.1; -
KW Hypothetical protein.
SQ SEQUENCE 388 AA; 43410 MW; A43F738BFD2E324A CRC64;

```

```

Query Match 50.6%; Score 45.5; DB 5; Length 388;
Best Local Similarity 60.0%; Pred. No. 19;
Matches 9; Conservative 2; Mismatches 3; Indels 1; Gaps 1;

QY 2 GYGRDPPDRPKKK 15
Db 338 GYGRDPPDRPKKK 352

```

```

RESULT 17
Q9FG28 PRELIMINARY; PRT; 209 AA.
AC Q9FG28;
DT 01-MAR-2001 (TREMBlrel. 16, Created)
DT 01-MAR-2001 (TREMBlrel. 16, Last sequence update)
DT 01-MAR-2001 (TREMBlrel. 16, Last annotation update)
DE Emb|CAB67623.1.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC Eucosids II; Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxId=3702;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=COLUMBIA;
RA Kaneko T., Katoh T., Asamizu E., Sato S., Nakamura Y., Kotani H.,
RA Tabata S.;
RT "Structural analysis of Arabidopsis thaliana chromosome 5. XI."
RL Submitted (MAY-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AP002032; BAB09813.1; -
SQ SEQUENCE 209 AA; 23892 MW; 9886AE479231CBFE CRC64;

```

```

Query Match 50.0%; Score 45; DB 10; Length 209;
Best Local Similarity 63.6%; Pred. No. 11;
Matches 7; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

QY 4 GRDNDPDRPSR 14
Db 138 GRDNDPDRPSQ 148

```

```
RESULT 18
ID Q93Q08 PRELIMINARY; PRT; 593 AA.
AC Q93Q08;
DT 01-DEC-2001 (TReMBLrel. 19, Created)
DT 01-DEC-2001 (TReMBLrel. 19, Last sequence update)
DT 01-DEC-2001 (TReMBLrel. 19, Last annotation update)
DE Aftmbrial adhesin Enf.
GN ENFA.
OS Escherichia coli.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
CX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=EMC 236;
RA Monteiro-Neto V., Giron J.A., Bando S.Y., Moreira-Filho C.A.,
RA Tschulsi L.R.;
RT "An afimbrial adhesin produced by enterocaggregative Escherichia coli
RT O11:H12."
RL Submitted (OCT-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF196977; AAK58509.1; -
SQ SEQUENCE 593 AA; 63025 MW; 4C396413A7B63C56 CRC64;

Query Match 50.0%; Score 45; DB 2; Length 593;
Best Local Similarity 57.1%; Pred. No. 37;
Matches 8; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

QY 3 YGRDNDRDPSPKRN 16
Db 520 YGRDQNNNTKRN 533

RESULT 19
ID Q9C5H5 PRELIMINARY; PRT; 716 AA.
AC Q9C5H5;
DT 01-JUN-2001 (TReMBLrel. 17, Created)
DT 01-JUN-2001 (TReMBLrel. 17, Last sequence update)
DT 01-MAR-2002 (TReMBLrel. 20, Last annotation update)
DE Putative MAP protein kinase.
DB MUD21.11/AT5G66850.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eucotsids II; Brassicales; Brassicaceae; Arabidopsis.
CX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RA Yamada K., Liu S.X., Sakano H., Pham P.K., Banh J., Chung M.K.,
RA Goldsmith A.D., Lee J.M., Quach H.L., Toriumi M., Yu G., Bowser L.,
RA Carninci P., Chen H., Cheuk R., Hayashizaki Y., Ishida U., Jones T.,
RA Kamiya A., Karlin-Neumann G., Kawai J., Kim C., Koesema E., Lam B., Lin J.,
RA Miranda M., Narusaka M., Nguyen M., Palm C.J., Sakurai T., Satou M.,
RA Seki M., Shim P., Southwick A., Shinzaki K., Davis R.W., Ecker J.R.,
RA Theologis A.;
RT "Full length cDNA of gene MUD21.11/AT5G66850 (GI:9758135).";
RL Submitted (MAR-2001) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RA Yamada K., Liu S.X., Sakano H., Pham P.K., Banh J., Chung M.K.,
RA Dale J.M., Gibson H.A., Goldsmith A.D., Jiang P.X., Lee J.M.,
RA Quach H.L., Tang C., Toriumi M., Yu G., Bowser L., Carninci P.,
RA Chen H., Cheuk R., Hayashizaki Y., Ishida U., Jones T., Kamiya A.,
RA Karlin-Neumann G., Kawai J., Kim C., Koesema E., Lam B., Lin J.,
RA Meyers M.C., Miranda M., Narusaka M., Nguyen M., Palm C.J.,
RA Sakurai T., Satou M., Seki M., Shim P., Southwick A., Tracy S.E.,
RA Shinozaki K., Davis R.W., Ecker J.R., Theologis A.;
RT "Full length cDNA of gene MUD21.11/AT5G66850 (GI:9758135).";
RL Submitted (JUN-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF360242; AAK25952.1; -
DR EMBL; AY040040; AAK64098.1; -
```

```
DR HSP, Q00534; 1817.
DR InterPro; IPR000719; Euk_pkinase.
DR InterPro; IPR002290; Ser_thr_pkinase.
DR InterPro; IPR001245; Tyr_pkinase.
DR Pfam; PF00069; Pkinase; 1.
DR PRINTS; PD00109; TYRKINASE.
DR ProDom; PD000001; Euk_pkinase; 1.
DR SMART; SM00220; S_TKc; 1.
DR SMART; SM00219; TYRKc; 1.
DR PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
DR PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
KW ATP-binding; Kinase; transferase.
SQ SEQUENCE 716 AA; 79319 MW; 9071054f2c6c68nd CRC64;

Query Match 50.0%; Score 45; DB 10; Length 716;
Best Local Similarity 72.7%; Pred. No. 46;
Matches 8; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 4 GRDNDRDPSPR 14
Db 124 GLDNRDRDPSPR 134

RESULT 20
ID Q93ZH4 PRELIMINARY; PRT; 716 AA.
AC Q93ZH4;
DT 01-DEC-2001 (TReMBLrel. 19, Created)
DT 01-DEC-2001 (TReMBLrel. 19, Last sequence update)
DT 01-DEC-2001 (TReMBLrel. 21, Last annotation update)
DE AT5G66850/MUD21_11.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eucotsids II; Brassicales; Brassicaceae; Arabidopsis.
CX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RA Cheuk R., Chen H., Kim C.J., Koesema E., Meyers M.C., Banh J.,
RA Bowser L., Carninci P., Dale J.M., Goldsmith A.D., Hayashizaki Y.,
RA Ishida U., Jiang P.X., Jones T., Kamiya A., Karlin-Neumann G.,
RA Kawai J., Lam B., Lee J.M., Lin J., Liu S.X., Miranda M., Narusaka M.,
RA Nguyen M., Onodera C.S., Palm C.J., Pham P.K., Quach H.L., Sakurai T.,
RA Satou M., Seki M., Southwick A., Tang C.C., Toriumi M., Yamada K.,
RA Yamamura Y., Yu G., Yu S., Shinozaki K., Davis R.W., Theologis A.,
RA Ecker J.R.;
RT "Arabidopsis cDNA clones.";
RL Submitted (SEP-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AY057533; AAL09773.1; -
DR InterPro; IPR000719; Euk_pkinase.
DR Pfam; PF00069; Pkinase; 1.
DR ProDom; PD000001; Euk_pkinase; 1.
DR PROSITE; PS00107; PROTEIN_KINASE_ATP; UNKNOWN_1.
DR PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
KW ATP-binding; transferase.
SQ SEQUENCE 716 AA; 79320 MW; 467039E12C6E271 CRC64;

Query Match 50.0%; Score 45; DB 10; Length 716;
Best Local Similarity 72.7%; Pred. No. 46;
Matches 8; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 4 GRDNDRDPSPR 14
Db 124 GLDNRDRDPSPR 134

RESULT 21
ID Q9ZWT3 PRELIMINARY; PRT; 810 AA.
AC Q9ZWT3;
DT 01-MAY-1999 (TReMBLrel. 10, Created)
DT 01-MAY-1999 (TReMBLrel. 10, Last sequence update)
DT 01-DEC-2001 (TReMBLrel. 19, Last annotation update)
```



```
DE PVL100. Cucurbita maxima (Pumpkin) (winter squash).
OS Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids I; Cucurbitales; Cucurbitaceae; Cucurbita.
OX NCBI_TaxID=3661;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=KUROKAWA AMAKURI NANKIN; TISSUE=COTYLEDON;
RC MEDLINE=99107919; PubMed=9891029;
RA Yamada K., Shimada T., Kondo M., Nishimura M., Hara-Nishimura I.,
RT "Multiple functional proteins are produced by cleaving Asn-Gln bonds
RL of a single precursor by vacuolar processing enzyme.",
DR J. Biol. Chem. 274:2563-2570(1999).
DR1 EMBL; AB019195; BAA34056.1; -.
DR HSSP; P02853; 2PHL.
DR InterPro; IPR001113; Seedstore_7s.
DR Pfam; PF00546; Seedstore_7s; 1.
DR Pfam; PF02808; Seedstore_7s; 1.
SQ SEQUENCE 810 AA; 97315 MW; A629A3F7542266AB CRC64;

Query March 50.0%; Score 45; DB 10; Length 810;
Best Local Similarity 63.6%; Pred. No. 52;
Matches 7; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Qy 5 RDDNDPSPSRK 15
||:|||||:
Db 158 RDDNERDPKRE 168

RESULT 22
Q95Q19 PRELIMINARY; PRT; 165 AA.
AC Q95Q19;
DT 01-DEC-2001 (TrEMBLrel. 13, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Hypothetical 17.3 kDa protein.
GN P25B5.7.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditidae;
OC Rhabditidae; Pelodidae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BRISTOL N2;
RC MEDLINE=99069613; PubMed=9851916;
RA None;
RT "Genome sequence of the nematode C. elegans: a platform for
RT investigating biology. The C. elegans Sequencing Consortium.",
RL Science 282:2012-2018(1998).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=BRISTOL N2;
RC STRAIN=BRISTOL N2;
RA Taich A.;
RT "The sequence of C. elegans cosmid P25B5.",
RL Submitted (APR-1995) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=BRISTOL N2;
RC STRAIN=BRISTOL N2;
RA Waterston R.;
RT "Direct Submission.",
RL Submitted (NOV-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; U23172; AL02495.1; -.
DR InterPro; IPR000504; RMA_rec_mot.
DR Pfam; PF00076; rtm; 1.
DR PROSITE; PSS0102; RRM; 1.
DR PROSITE; PSS0030; RRM_RNP_1; UNKNOWN_1.
KM Hypothetical protein.
SQ SEQUENCE 165 AA; 17316 MW; 077B146427190BFB CRC64;

Query Match 48.9%; Score 44; DB 5; Length 165;
Best Local Similarity 47.6%; Pred. No. 13;

Matches 10; Conservative 1; Mismatches 2; Indels 8; Gaps 1;

Qy 4 GRDD-----NDRDPSRKN 16
|||||
Db 13 GRDDSSRYAKAKENDRKPERN 33

RESULT 23
Q9R6A3 PRELIMINARY; PRT; 394 AA.
AC Q9R6A3;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE TIRP174 protein.
DR TIRP174.
GN TIRP174.
OS Agrobacterium tumefaciens.
OC Plasmid pTi-SAKURA.
OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
OC Rhizobiaceae; Rhizobium.
OX NCBI_TaxID=358;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MAFF301001; PubMed=10721727;
RC MEDLINE=20184752; PubMed=10721727;
RA Suzuki K., Hattori Y., Uraji M., Ohta N., Iwata K., Murata K.,
RA Katoh A., Yoshida K.;
RT "Complete nucleotide sequence of a plant tumor-inducing Ti plasmid.",
RL Gene 242:331-336(2000).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=MAFF301001; PubMed=9524202;
RC MEDLINE=98193120; PubMed=9524202;
RA Suzuki K., Ohta N., Hattori Y., Uraji M., Katoh A., Yoshida K.;
RT "Novel structural difference between nopaline- and octopine-type trbJ
RT gene: construction of genetic and physical map and sequencing of
RT trbJ/trai and rep gene clusters of a new Ti plasmid pTi-SAKURA.",
RL Biochim. Biophys. Acta 1396:1-7(1998).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=MAFF301001;
RA Hattori Y., Suzuki K., Ohta N., Uraji M., Katoh A., Yoshida K.;
RT "Genome structure of pTi-SAKURA(I): Strategy for DNA sequencing of a
RT Japanese cherry-Ti plasmid.",
RL Nucleic Acids Symp. Ser. 37:159-160(1998).
RN [4]
RP SEQUENCE FROM N.A.
RC STRAIN=MAFF301001;
RA Ohta N., Suzuki K., Hattori Y., Uraji M., Katoh A., Yoshida K.;
RT "Genome structure of pTi-SAKURA (III): Characteristics of T-DNA.",
RL Nucleic Acids Symp. Ser. 39:185-186(1998).
RN [5]
RP SEQUENCE FROM N.A.
RC STRAIN=MAFF301001;
RA Uraji M., Suzuki K., Ohta N., Hattori Y., Katoh A., Yoshida K.;
RT "Genome structure of pTi-SAKURA (IV): Characteristics of tra region.",
RL Nucleic Acids Symp. Ser. 39:187-188(1998).
RN [6]
RP SEQUENCE FROM N.A.
RC STRAIN=MAFF301001;
RA Hattori Y., Suzuki K., Ohta N., Uraji M., Katoh A., Yoshida K.;
RT "Genome structure of pTi-SAKURA (V): Complete nucleotide sequence of
RT plasmid pTi-SAKURA's vir region in Agrobacterium tumefaciens.",
RL Nucleic Acids Symp. Ser. 39:265-266(1998).
DR EMBL; AB016260; BAA87799.1; -.
KM Plasmid.
SQ SEQUENCE 394 AA; 44533 MW; 613B5B8A18AFD20A CRC64;

Query Match 48.9%; Score 44; DB 2; Length 394;
Best Local Similarity 66.7%; Pred. No. 34;
Matches 8; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Qy 1 VGYGRDNDPDRP 12
```

Db 249 VGH1ADDNRNP 260

## RESULT 24

Q9LNM5 PRELIMINARY; PRT; 443 AA.  
 DT 01-OCT-2000 (TREMBlurel. 15, Created)  
 DT 01-OCT-2000 (TREMBlurel. 15, Last sequence update)  
 DT 01-JUN-2002 (TREMBlurel. 21, Last annotation update)  
 DE F16F4.11 protein (Hypothetical 50.9 kDa protein).  
 GN F16F4.11 OR AT1G21200.  
 OS Arabidopsis thaliana (Mouse-ear cress).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;  
 OC Eustoids II; Brassicales; Brassicaceae; Arabidopsis.  
 NCBI\_TaxID=3702;  
 RN  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=CV. COLUMBIA;  
 RA Liu S.X., Sakano H., Yu G., Ergu P., Lee J.M., Lenz C., Pham P.,  
 RA Toriumi M., Byun E., Chan A., Chin C., Choi E., Chung M.,  
 RA Goldsmith A., Gonzalez A., Liu A., Smith A., Vaysberg M., Altafi H.,  
 RA Brooks S., Buehler E., Chao Q., Conn L., Conway A.B., Hansen N.F.,  
 RA Johnson-Hopson C., Khan S., Kim C., Lam B., Miranda M., Nguyen M.,  
 RA Palm C.J., Shinn P., Southwick A., Davis R.W., Ecker J.R.,  
 RA Federpiehl N.A., Theologis A.,  
 RT "The sequence of BAC F16F4 from Arabidopsis thaliana chromosome 1.",  
 RL Submitted (JUL-2000) to the EMBL/GenBank/DBJ databases.  
 RN  
 RP SEQUENCE FROM N.A.  
 RA Yamada K., Liu S.X., Pham P.K., Banh J., Dale J.M., Goldsmith A.D.,  
 RA Jiang P.X., Lee J.M., Onodera C.S., Quach H.L., Tang C., Toriumi M.,  
 RA Yamamura Y., Yu G., Yu S., Bowser L., Carninci P., Chen H., Cheuk R.,  
 RA Hayashizaki Y., Ishida J., Jones T., Kamita A., Karlin-Neumann G.,  
 RA Kawai J., Kim C., Koeseema E., Lam B., Lin J., Meyers M.C., Miranda M.,  
 RA Natusaka M., Nguyen M., Palm C.J., Sakurai T., Satou M., Seki M.,  
 RA Shinn P., Southwick A., Tracy S.E., Shinzaki K., Davis R.W.,  
 RA Ecker J.R., Theologis A.,  
 RT "Full length cDNA of gene F16F4.11 (GI:8920640).",  
 RL Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.  
 RN  
 RP SEQUENCE FROM N.A.  
 RA Yamada K., Banh J., Chan M.M., Chang C.H., Chang E., Dale J.M.,  
 RA Deng J.M., Goldsmith A.D., Lee J.M., Onodera C.S., Quach H.L.,  
 RA Tang C., Toriumi M., Wu H.C., Yamamura Y., Yu G., Bowser L.,  
 RA Carninci P., Chen H., Cheuk R., Hayashizaki Y., Ishida J., Jones T.,  
 RA Kamita A., Karlin-Neumann G., Kawai J., Kim C., Lam B., Lin J.,  
 RA Meyers M.C., Miranda M., Natusaka M., Nguyen M., Palm C.J.,  
 RA Sakurai T., Satou M., Seki M., Shinn P., Southwick A., Shinzaki K.,  
 RA Davis R.W., Ecker J.R., Theologis A.,  
 RT "Arabidopsis Open Reading Frame (ORF) Clones.",  
 RL Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AC036104; AAF81362.1; -;  
 DR EMBL; AY045850; AAK76524.1; -;  
 DR EMBL; AY091378; AAM14317.1; -;  
 DR InterPro; IPR001005; MYD\_DNA\_binding.  
 DR SMART; SMO0395; SANT; 1.  
 KW Hypothetical protein.  
 SQ SEQUENCE 443 AA; 50933 MW; F920D3BF3C95057B CRC64;

Query Match 48.9%; Score 44; DB 10; Length 443;  
 Best Local Similarity 66.7%; Pred. No. 39;  
 Matches 8; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 5 RDDNDPDRSKN 16  
 DB 265 RDDNDPDRSKN 276

RESULT 25  
 Q8TIN3

ID Q8TIN3 PRELIMINARY; PRT; 515 AA.  
 AC Q8TIN3;  
 DT 01-JUN-2002 (TREMBlurel. 21, Created)  
 DT 01-JUN-2002 (TREMBlurel. 21, Last sequence update)  
 DT 01-JUN-2002 (TREMBlurel. 21, Last annotation update)  
 DE Hypothetical 61.5 kDa protein.  
 OS Dictyostelium discoideum (Slime mold).  
 OC Eukaryota; Mycetozoa; Dictyostelida; Dictyostelium.  
 NCBI\_TaxID=44689;  
 RN  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=AX4;  
 RA Gloeckner G., Eichinger L., Szafranski K., Pachebat J., Dear P.,  
 RA Lehmann R., Baumgart C., Parra G., April J.F., Guigo R., Kumpf K.,  
 RA Tunggal B., Cox E., Quail M.A., Platzer M., Rosenthal A., Noegel A.A.,  
 RT "Sequence and Analysis of Chromosome 2 of Dictyostelium.",  
 RL Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AC116350; AAL96720.1; -;  
 KW Hypothetical protein.  
 SQ SEQUENCE 515 AA; 61503 MW; 2761DE2105859E9F CRC64;

Query Match 48.9%; Score 44; DB 5; Length 515;  
 Best Local Similarity 69.2%; Pred. No. 46;  
 Matches 9; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 3 YGRDNDPDRSKN 15  
 DB 328 YKIDNDPDRSTK 340

## RESULT 26

Q09542 PRELIMINARY; PRT; 561 AA.  
 AC Q09542;  
 DT 01-NOV-1996 (TREMBlurel. 01, Created)  
 DT 01-NOV-1996 (TREMBlurel. 01, Last sequence update)  
 DT 01-JUN-2002 (TREMBlurel. 21, Last annotation update)  
 DE Hypothetical 60.3 kDa protein.  
 GN F25B5.7.  
 OS Caenorhabditis elegans.  
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;  
 OC Rhabditidae; Pelodermidae; Caenorhabditis.  
 NCBI\_TaxID=6239;  
 RN  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=BRISTOL N2;  
 RX MEDLINE=99069613; PubMed=9851916;  
 RA None;  
 RT "Genome sequence of the nematode C. elegans: a platform for  
 RT investigating biology. The C. elegans Sequencing Consortium.",  
 RL Science 282:2012-2018 (1998).  
 RN  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=BRISTOL N2;  
 RA Tach A.,  
 RT "The sequence of C. elegans cosmid F25B5.",  
 RL Submitted (APR-1995) to the EMBL/GenBank/DBJ databases.  
 RN  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=BRISTOL N2;  
 RA Waterston R.,  
 RT "Direct Submission.",  
 RL Submitted (SEP-2001) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; U23172; AAC46528.1; -;  
 DR InterPro; IPR000504; RNA\_rec\_mot.  
 DR Pfam; PF00076; rrm; 2.  
 DR SMART; SMO0360; RRM; 2.  
 DR PROSITE; PS0102; RRM; 2.  
 DR PROSITE; PS00030; RRM\_RNP\_1; UNKNOWN\_1.  
 KW Hypothetical protein.  
 SQ SEQUENCE 561 AA; 60285 MW; 22DBFC8C9861E029 CRC64;

Query Match 48.9%; Score 44; DB 5; Length 561;

Best Local Similarity 47.6%; Pred. No. 51;  
Matches 10; Conservative 1; Mismatches 2; Indels 8; Gaps 1;  
QY 4 GRDD-----NDRDPSRKN 16  
DB 13 GRDSSRYAKAENDRKPGREN 33

RESULT 27  
096554 PRELIMINARY; PRT; 8243 AA.  
AC 096554;  
DT 01-MAY-1999 (TREMBlrel. 10, Created)  
DT 01-MAY-1999 (TREMBlrel. 10, Last sequence update)  
DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)  
DE TYPE I fatty acid synthase.  
GN PAS1.  
OS Cryptosporidium parvum.  
OC Eukaryota; Alveolata; Apicomplexa; Coccidia; Eimeriida;  
OC Cryptosporidiidae; Cryptosporidium.  
OX NCBI\_Taxid=5807;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=KCU-1;  
RX MEDLINE=20155414; PubMed=10693747;  
RA Zhu G., Marchewka M.J., Woods K.M., Upton S.J., Keitely J.S.;  
RT "Molecular analysis of a Type I fatty acid synthase in Cryptosporidium  
parvum."  
RL Mol. Biochem. Parasitol. 105:253-260(2000).  
DR EMBL; AF082993; AAC9407.1; -.  
DR HSSP; P28304; 10OR.  
DR InterPro; IPR001227; Ac transferase.  
DR InterPro; IPR002085; Adh zn family.  
DR InterPro; IPR000873; AMP-bind.  
DR InterPro; IPR000794; Ketoacyl-synt.  
DR InterPro; IPR003880; Prantne attach.  
DR Pfam; PF00698; Acyl\_transf; 3.  
DR Pfam; PF00107; adh\_zinc; 3.  
DR Pfam; PF00501; AMP-binding; 2.  
DR Pfam; PF00109; ketoacyl-synt; 3.  
DR Pfam; PF02801; ketoacyl-synt\_C; 3.  
DR Pfam; PF00550; pp-binding; 4.  
DR PROSITE; PS50075; ACP\_DOMAIN; 4.  
DR PROSITE; PS50045; ACP\_BINDING; 4.  
DR PROSITE; PS00606; B\_KETOACYL\_SYNTHASE; UNKNOWN\_1.  
DR PROSITE; PS00012; PHOSPHOPANTHETHEINE; UNKNOWN\_2.  
KM Phosphopantetheine.  
SQ SEQUENCE 8243 AA; 920806 MW; F5A1BC8E4B3E3E7C CRC64;  
Query Match 48.9%; Score 44; DB 5; Length 8243;  
Best Local Similarity 57.1%; Pred. No. 1; Le+03;  
Matches 8; Conservative 1; Mismatches 5; Indels 0; Gaps 0;  
QY 3 YGRDDNDRDPSPRKN 16  
DB 4317 YGRDDDELPPSPSN 4330  
RESULT 28  
094HP4 PRELIMINARY; PRT; 139 AA.  
AC 094HP4;  
DT 01-DEC-2001 (TREMBlrel. 19, Created)  
DT 01-DEC-2001 (TREMBlrel. 19, Last sequence update)  
DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)  
DE Hypothetical protein.  
OS Oryza sativa (Rice).  
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;  
OC Ehrhartoideae; Oryzaceae; Oryza.  
OX NCBI\_Taxid=4530;  
RN [1]  
RP SEQUENCE FROM N.A.

RC STRAIN=NIPPOBARE;  
RA Spiegel L., de la Bastide M., Nascimento L., Kirchoff K., King L.,  
RA Preston R., Vil M.D., Baker J., Bell M., Zutavern T., Santos L.,  
RA Miller B., Kuit K., Rodriguez S., Cummins D.M., Balija V., Shah R.,  
RA Bahret A., Bal H., O'Shaughnessy A., Dedhia N., McCombie W.R.;  
RT "Genomic Sequence for Oryza sativa, Nipponbare strain, clone  
OSUNBa0089D15, from Chromosome 10, complete sequence."  
RT Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.  
RL [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=NIPPOBARE;  
RA McCombie W.R.;  
RL Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.  
RN [3]  
RP SEQUENCE FROM N.A.  
RC STRAIN=NIPPOBARE;  
RA McCombie W.R.;  
RL Submitted (MAY-2001) to the EMBL/GenBank/DBJ databases.  
RN [4]  
RP SEQUENCE FROM N.A.  
RC STRAIN=NIPPOBARE;  
RA Palmer L.E., Spiegel L., de la Bastide M., Nascimento L., Kirchoff K.,  
RA King L., Preston R., Vil M.D., Baker J., Bell M., Zutavern T.,  
RA Santos L., Miller B., Kuit K., Rodriguez S., Cummins D.M., Balija V.,  
RA Shah R., Bahret A., Bal H., O'Shaughnessy A., Dedhia N.,  
RA McCombie W.R.;  
RL Submitted (AUG-2001) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AC078944; AAK92593.1; -. 78F547B57D4B0B45 CRC64;  
SQ SEQUENCE 139 AA; 14962 MW; 78F547B57D4B0B45 CRC64;  
Query Match 47.8%; Score 43; DB 10; Length 139;  
Best Local Similarity 53.8%; Pred. No. 15;  
Matches 7; Conservative 2; Mismatches 4; Indels 0; Gaps 0;  
QY 2 GYGRDDNDRDPSPR 14  
DB 127 GYGRDDNDPSDEPR 139  
RESULT 29  
0919J2 PRELIMINARY; PRT; 142 AA.  
AC 0919J2;  
DT 01-OCT-2000 (TREMBlrel. 15, Created)  
DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)  
DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)  
DE P75 neurotrophin receptor b (fragment).  
GN P75NTRB.  
OS Xenopus laevis (African clawed frog).  
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipidoidea; Pipidae;  
OC Xenopodidae; Xenopus.  
OX NCBI\_Taxid=8355;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Hutson L.D., Bothwell M.;  
RT "Developmental expression of two putative p75NTR genes in Xenopus  
laevis."  
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AF246463; AAF63523.1; -.  
DR HSSP; O14763; 1DOG.  
DR InterPro; IPR001368; TNFR\_c6.  
DR Pfam; PF00020; TNFR\_c6; 2.  
DR SMART; SM00208; TNFR; 2.  
DR PROSITE; PS00652; TNFR\_NGFR\_1; UNKNOWN\_1.  
DR PROSITE; PS50050; TNFR\_NGFR\_2; 1.  
KM Receptor.  
FT NON\_TER 1 1  
FT NON\_TER 142 142  
SQ SEQUENCE 142 AA; 15120 MW; 3C50CFE7DD1B57FD CRC64;  
Query Match 47.8%; Score 43; DB 13; Length 142;  
Best Local Similarity 70.0%; Pred. No. 16;

```
Matches 7, Conservative 1, Mismatches 2, Indels 0, Gaps 0,
Qy 3 YGRDDNDRDP 12
|:|||||
Db 40 YSDEDNDRDP 49

RESULT 30
ID 0919J3 PRELIMINARY; PRT; 143 AA.
AC 0919J3;
DT 01-OCT-2000 (TREMBlrel. 15, Created)
DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)
DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
DE p75 neurotrophin receptor a (Fragment).
GN P75NTRA.
OS Xenopus laevis (African clawed frog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipidea; Pipidae;
OC Xenopodinae; Xenopus.
OX NCBI_TaxID=8355;
RN [1]
RP SEQUENCE FROM N.A.
RA Hutson L.D., Bothwell M.;
RT "Developmental expression of two putative p75NTR genes in Xenopus
laevis."
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF246462; AAF63522.1; -.
DR HSSP; P19438; IEXT.
DR InterPro; IPR001899; Gram_pos_anchor.
DR InterPro; IPR001368; TNFR_c6.
DR Pfam; PF00020; TNFR_c6; 2.
DR SMART; SM00208; TNFR; 2.
DR PROSITE; PS00343; GRAM_POS_ANCHORING; UNKNOWN_1.
DR PROSITE; PS00652; TNFR_NGFR_1; UNKNOWN_1.
DR PROSITE; PS50050; TNFR_NGFR_2; 1.
DR Receptor.
FT NON_TER 1
FT NON_TER 143
SQ SEQUENCE 143 AA; 15031 MW; D583924B847AC324 CRC64;

Query Match 47.8%; Score 43; DB 13; Length 143;
Best Local Similarity 70.0%; Pred. No. 16;
Matches 7, Conservative 1, Mismatches 2, Indels 0, Gaps 0;
Qy 3 YGRDDNDRDP 12
|:|||||
Db 40 YSDEDNDRDP 49

RESULT 31
ID 09FIE4 PRELIMINARY; PRT; 219 AA.
AC 09FIE4;
DT 01-MAR-2001 (TREMBlrel. 16, Created)
DT 01-MAR-2001 (TREMBlrel. 16, Last sequence update)
DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
DE Genomic DNA, chromosome 5, p1 clone:MSF19 (Hypothetical 26.0 kda
protein).
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eustosids II; Brassicales; Brassicaceae; Arabidopsids.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=COLUMBIA;
RX MEDLINE=99156233; PubMed=10048488;
RA Asamizu E., Sato S., Kaneko T., Nakamura Y., Kotani H., Miyajima N.,
RA Tabata S.;
RT "Structural analysis of Arabidopsis thaliana chromosome 5. VIII.
Sequence features of the regions of 1,081,958 bp covered by seventeen
physically assigned P1 and TAC clones.";
```

```
RL DNA Res. 5:379-391(1998).
RN [2]
RP SEQUENCE FROM N.A.
RA Yamada K., Liu S.X., Sakano H., Pham P.K., Banh J., Chung M.K.,
RA Goldsmith A.D., Lee J.M., Quach H.L., Tang C.C., Toriumi M., Yu G.,
RA Bowser L., Carinici P., Chen H., Cheuk R., Hayashizaki Y., Ishida J.,
RA Jones T., Kamiya A., Karlin-Neumann G., Kawai J., Kim C., Koessma B.,
RA Lam B., Lin J., Meyers M.C., Miranda M., Narusaka M., Nguyen M.,
RA Palm C.J., Sakurai T., Satou M., Seki M., Shim P., Southwick A.,
RA Shinozaki K., Davis R.W., Ecker J.R., Theologis A.;
RT "Full length cDNA of gene MSF19.3/AT5G57370 (GI:9757931).";
RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AB016891; BAB08474.1; -.
DR EMBL; AF370233; AAK44048.1; -.
KW Hypothetical protein.
SQ SEQUENCE 219 AA; 26041 MW; D7577F3B135C872C CRC64;

Query Match 47.8%; Score 43; DB 10; Length 219;
Best Local Similarity 72.7%; Pred. No. 26;
Matches 8, Conservative 0, Mismatches 3, Indels 0, Gaps 0;
Qy 4 GRDDNDRDPSR 14
|:|||||
Db 24 GRDDNDRDPSR 34

RESULT 32
ID 09PRG7 PRELIMINARY; PRT; 401 AA.
AC 09PRG7;
DT 01-MAY-2000 (TREMBlrel. 13, Created)
DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)
DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)
DE p75 neurotrophin receptor A-2.
GN P75NTRA.
OS Xenopus laevis (African clawed frog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipidea; Pipidae;
OC Xenopodinae; Xenopus.
OX NCBI_TaxID=8355;
RN [1]
RP SEQUENCE FROM N.A.
RA Hutson L.D., Richards A.P., Bothwell M.;
RT "Life and death in the developing Xenopus retina: The role of the p75
neurotrophin receptor."
RL Submitted (JUL-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF172400; AAD51031.1; -.
DR EMBL; AF172399; AAD51030.1; -.
DR HSSP; P07174; INGR.
DR InterPro; IPR000488; Death.
DR InterPro; IPR000561; EGF-like.
DR InterPro; IPR001368; TNFR_c6.
DR Pfam; PF00531; death; 1.
DR Pfam; PF00020; TNFR_c6; 4.
DR SMART; SM00005; DEATH; 1.
DR SMART; SM00208; TNFR; 4.
DR PROSITE; PS50017; DEATH_DOMAIN; 1.
DR PROSITE; PS01186; EGF_2; UNKNOWN_1.
DR PROSITE; PS00652; TNFR_NGFR_1; UNKNOWN_3.
DR PROSITE; PS50050; TNFR_NGFR_2; 3.
DR Receptor.
SQ SEQUENCE 401 AA; 43419 MW; D13BCAF7863FEFCF CRC64;

Query Match 47.8%; Score 43; DB 13; Length 401;
Best Local Similarity 70.0%; Pred. No. 51;
Matches 7, Conservative 1, Mismatches 2, Indels 0, Gaps 0;
Qy 3 YGRDDNDRDP 12
|:|||||
Db 144 YSDEDNDRDP 153

RESULT 33
```

Q90299 ID Q90299 PRELIMINARY; PRT; 414 AA.  
 AC Q90299;  
 DT 01-DEC-2001 (TREMBLrel. 19, Created)  
 DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)  
 DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)  
 DE PCDD9.  
 GN PCDD9.  
 OS Xenopus laevis (African clawed frog).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipridae; Pipidae;  
 OC Xenopodinae; Xenopus.  
 NCBI\_TaxID=8355;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Cardeguas J.A., Bogenhagen D.F.;  
 RT "The Xenopus homolog of chicken pro-apoptotic protein p52 and human  
 RT PCDD9 is a mitochondrial protein."  
 RL Submitted (OCT-2000) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AF317657; AAK6969.1; -  
 SO SEQUENCE 414 AA; 47229 MW; 45B81A4C9EDFCAB6 CRC64;

Query Match 47.8%; Score 43; DB 13; Length 414;  
 Best Local Similarity 70.0%; Pred. No. 53;  
 Matches 7; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

Oy 7 DNDSDPSRKN 16  
 Db 366 DSDKPSRKN 375

## RESULT 34

006860 ID 006860 PRELIMINARY; PRT; 419 AA.  
 AC 006860; 009365;  
 DT 01-JUL-1997 (TREMBLrel. 04, Created)  
 DT 01-JUL-1997 (TREMBLrel. 04, Last sequence update)  
 DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)  
 DE Gyrase beta subunit (EC 5.99.1.3) (Fragment).  
 GN GYRB.  
 OS Shewanella putrefaciens.  
 OC Bacteria; Proteobacteria; gamma subdivision; Alteromonadales;  
 OC Shewanella.  
 NCBI\_TaxID=24;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=IM-1;  
 RL Venkateswaran K., Neilson K.H.;  
 RL Submitted (MAY-1997) to the EMBL/GenBank/DBJ databases.  
 CC -1- FUNCTION: DNA GYRASE NEGATIVELY SUPERCOILS CLOSED CIRCULAR DOUBLE-  
 CC STRANDED DNA IN AN ATP-DEPENDANT MANNER AND ALSO CATALYZES THE  
 CC INTERCONVERSION OF OTHER TOPOLOGICAL ISOMERS OF DOUBLE-STRANDED  
 CC DNA RINGS, INCLUDING CATENANES AND KNOTTED RINGS.  
 CC -1- CATALYTIC ACTIVITY: ATP-DEPENDENT BREAKAGE, PASSAGE AND REJOINING  
 CC OF DOUBLE-STRANDED DNA.  
 CC -1- SUBUNIT: MADE UP OF TWO CHAINS. THE A CHAIN IS RESPONSIBLE FOR DNA  
 CC BREAKAGE AND REJOINING, THE B CHAIN CATALYZES ATP HYDROLYSIS. THE  
 CC ENZYME FORMS AN A2B2 TETRAMER.  
 DR EMBL; AF005681; AAB80824.1; -  
 DR HSSP; P06982; 1A96.  
 DR InterPro; IPR001241; DNA\_topoisom1.  
 DR Pfam; PF00204; DNA\_gyraseB; 1.  
 DR PRINTS; PR00418; TP22FAMILY.  
 DR SMART; SM00433; TOP2c; 1.  
 KW Isomerase; Topoisomerase; ATP-binding.  
 FT NON TER 1  
 FT NON TER 1  
 SO SEQUENCE 419 AA; 46654 MW; E4959C5F1910C257 CRC64;

Query Match 47.8%; Score 43; DB 2; Length 419;  
 Best Local Similarity 61.5%; Pred. No. 54;  
 Matches 8; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

Oy 2 GYGRDNDPSR 14  
 Db 386 GIGEDDHSASR 398

## RESULT 35

09VD47 ID 09VD47 PRELIMINARY; PRT; 444 AA.  
 AC 09VD47;  
 DT 01-MAY-2000 (TREMBLrel. 13, Created)  
 DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)  
 DT 01-JUN-2000 (TREMBLrel. 14, Last annotation update)  
 DE CG13407 protein.  
 GN CG13407.  
 OS Drosophila melanogaster (Fruit fly).  
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
 OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
 OC Ephydriidae; Drosophilidae; Drosophila.  
 NCBI\_TaxID=7227;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=BERKELEY;  
 RX MEDLINE=20196006; PubMed=10731132;  
 RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,  
 RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,  
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,  
 RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,  
 RA Brandon R.C., Rogers Y.-H.C., Blaise R.G., Champe M., Pfeiffer B.D.,  
 RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,  
 RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,  
 RA Ballew R.M., Baas A., Baxendale J., Bayraktaroglu L., Beasley E.M.,  
 RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,  
 RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brotler P.,  
 RA Burris K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,  
 RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,  
 RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,  
 RA Dodson K., Doup I.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,  
 RA Dushin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,  
 RA Fowler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,  
 RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,  
 RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,  
 RA Houston D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,  
 RA Jaiswal B., Kalish F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,  
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,  
 RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang X., Lin X.,  
 RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,  
 RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,  
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,  
 RA Nelson D.R., Nelson K.A., Nixon K., Nuskern D.R., Paclob J.M.,  
 RA Palazolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,  
 RA Reinert K., Remington K., Sanders R.D.C., Scheeler F., Shen H.,  
 RA Shue B.C., Siden-Kiamos I., Simpson W., Skipski M.P., Smith T.,  
 RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,  
 RA Wang Z.-Y., Wassarman D.A., Weinstein G.M., Weissenbach J.,  
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,  
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,  
 RA Zhang X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,  
 RA Gbbs R.A., Myers E.W., Rubin G.M., Venter J.C.;  
 RT "The genome sequence of Drosophila melanogaster";  
 RL Science 287:2185-2195(2000).  
 DR EMBL; AB003737; AAF55956.1; -  
 DR FlyBase; FBgn0038931; CG13407.  
 SO SEQUENCE 444 AA; 49232 MW; 0D6173811930592 CRC64;

Query Match 47.8%; Score 43; DB 5; Length 444;  
 Best Local Similarity 61.5%; Pred. No. 58;  
 Matches 8; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Oy 2 GYGRDNDPSR 14  
 Db 402 GYGRDNDPSR 414

```

RESULT 36
Q969S3      PRELIMINARY;      PRT;      477 AA.
ID 0969S3;
AC 0969S3;
DT 01-DEC-2001 (TRENBLREL. 19, Created)
DT 01-DEC-2001 (TRENBLREL. 19, Last sequence update)
DT 01-JUN-2002 (TRENBLREL. 21, Last annotation update)
DE Zinc finger-like protein 9 (Unknown) (Protein for MGC:2485) (Protein
for MGC:17552).
GN ZPR9.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Seong H.A., Gil M., Kim S.J., Kim K.T., Ha H.;
RT "Phosphorylation of ZPR9, a novel zinc finger-like protein, by MPK38
RT serine/threonine kinase.";
RL Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RA TISSUE=SKIN;
RC Strauberg R.;
RA Strauberg R.;
RL Submitted (MAY-2001) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC TISSUE=CERVIX;
RA Strauberg R.;
RL Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AY046059; AAL02121.1; -
DR EMBL; BC008752; AAH08752.1; -
DR EMBL; BC010545; AAH010545.1; -
DR InterPro: IPR000822; Znf_C2H2;
DR SMART: SM00355; Znf_C2H2; 4.
DR PROSITE; PS00028; ZINC_FINGER_C2H2_1; UNKNOWN_2.
KW DNA-binding; Zinc-finger.
SQ SEQUENCE 477 AA; 54272 MW; 88E75FAEAB46C628 CRC64;

Query Match      47.8%; Score 43; DB 4; Length 477;
Best Local Similarity 66.7%; Pred. NO. 62;
Matches 8; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY      4 GRDDNDPDRSRK 15
DB      165 GRGTHDRDPSEK 176

RESULT 37
Q9WVT7      PRELIMINARY;      PRT;      523 AA.
ID 09WVT7;
AC 09WVT7;
DT 01-NOV-1999 (TRENBLREL. 12, Created)
DT 01-NOV-1999 (TRENBLREL. 12, Last sequence update)
DT 01-DEC-2001 (TRENBLREL. 19, Last annotation update)
DE Hypothetical 59.1 kDa protein.
GN TORF8 OR C.
OS Agrobacterium tumefaciens (strain C58 / ATCC 33970).
OC Plasmid pTIC58.
OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
OC Rhizobiaceae; Rhizobium.
OX NCBI_TaxID=176299;
RN [1]
RP SEQUENCE FROM N.A.
RA STRAIN=C58;
RA Gielen J., Terry N., van Montagu M., Villarroel R.;
RT "Complete nucleotide sequence of the T-DNA region of the plant tumor
RT inducing Agrobacterium tumefaciens Ti plasmid pTIC58.";
RL Submitted (MAR-1999) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RA STRAIN=C58;

```

```

RA Otten L., de Ruifray P., Hammann P.;
RT "Left part of the C58 T-DNA.";
RL Submitted (FEB-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AJ237588; CAB44632.1; -
DR EMBL; AF126445; AAD30483.1; -
KW Hypothetical protein; Plasmid.
SQ SEQUENCE 523 AA; 59072 MW; 7605923629B51608 CRC64;

Query Match      47.8%; Score 43; DB 2; Length 523;
Best Local Similarity 66.7%; Pred. NO. 69;
Matches 8; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY      1 VGYGRDNDPDRP 12
DB      378 VGH1ADNDPDRSP 389

RESULT 38
Q8U6A8      PRELIMINARY;      PRT;      536 AA.
ID 08U6A8;
AC 08U6A8;
DT 01-JUN-2002 (TRENBLREL. 21, Created)
DT 01-JUN-2002 (TRENBLREL. 21, Last sequence update)
DT 01-JUN-2002 (TRENBLREL. 21, Last annotation update)
DE C protein.
GN GENE C OR ATU6002 OR AGR_PTI_41.
OS Agrobacterium tumefaciens (strain C58 / ATCC 33970).
OC Plasmid pTIC58.
OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
OC Rhizobiaceae; Rhizobium.
OX NCBI_TaxID=176299;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21608550; PubMed=11743193;
RA Wood D.W., Setubal J.C., Kaul R., Monks D.E., Kiraajima J.P.,
RA Okura Y., Zhou Y., Chen L., Wood G.E., Almeida N.F., Jr., Woo L.,
RA Chou Y., Paulsen I.T., Eisen J.A., Karp P.D., Boyee D., Sr.,
RA Chapman P., Clendenning J., Deatherage G., Gillet M., Grant C.,
RA Kuyavayn T., Levy R., Li M.-J., McClelland E., Palmeri A.,
RA Raymond C., Rouse G., Saenphimachak C., Wu Z., Romero P., Gordon D.,
RA Zhang S., Yoo H., Tao Y., Biddle P., Jung M., Krespan W., Perry M.,
RA Gordon-Kamm B., Liao L., Kim S., Hendrick C., Zhao Z.-Y., Dolan M.,
RA Chumley F., Tingey S.V., Tomb J.-F., Gordon M.P., Olson M.V.,
RA Neester E.W.;
RT "The genome of the natural genetic engineer Agrobacterium tumefaciens
RT C58.";
RL Science 294:2317-2323(2001).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=21608551; PubMed=11743194;
RA Goodner B., Hinkle G., Gattung S., Miller N., Blanchard M.,
RA Ourcello B., Goldman B.S., Cao Y., Askenazi M., Halling C., Mullin L.,
RA Houmiel K., Gordon J., Vaudin M., Iartchouk O., Epp A., Liu F.,
RA Flanagan C., Crowell C., Gurtson J., Lomo C., Sear C., Strub G.,
RA Cielo C., Slater S.;
RT "Genome sequence of the plant pathogen and biotechnology agent
RT Agrobacterium tumefaciens C58.";
RL Science 294:2323-2328(2001).
DR EMBL; AE009418; AAL46238.1; -
DR EMBL; AE007926; AAK90961.1; -
KW Plasmid; Complete proteome.
SQ SEQUENCE 536 AA; 60403 MW; 4FBECD9D7E594D30 CRC64;

Query Match      47.8%; Score 43; DB 16; Length 536;
Best Local Similarity 66.7%; Pred. NO. 71;
Matches 8; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY      1 VGYGRDNDPDRP 12
DB      391 VGH1ADNDPDRSP 402

```

```

RESULT 39
ID 024041 PRELIMINARY; PRT; 570 AA.
AC 024041;
DT 01-JAN-1998 (TREMblrel. 05, Created)
DT 01-JAN-1998 (TREMblrel. 05, Last sequence update)
DE 01-DEC-2001 (TREMblrel. 19, Last annotation update)
GN Laccase (EC 1.10.3.2).
OS Litodendron tulipifera.
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Magnoliales; Magnoliaceae; Litodendron.
OX NCBI_TaxID=3415;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=LIGNIFYING XYLEM; PubMed=10394942;
RX MEDLINE=99320865;
RA LaFayette P.R., Eriksson K.E., Dean J.F.D.;
RT "Characterization and heterologous expression of laccase cDNAs from
  xylem tissues of yellow-poplar (Litodendron tulipifera).";
RL Plant Mol. Biol. 40:23-35(1999).
DR EMBL; U73103; AAB17191.1; -.
DR HSP; P37064; IAOZ.
DR InterPro; IPR001117; Cu-oxidase.
DR InterPro; IPR002355; MultiCu_oxidase2.
DR Pfam; PF00394; Cu-oxidase; 3.
DR PROSITE; PS00079; MULTICOPPER_OXIDASE1; UNKNOWN_1.
DR PROSITE; PS00080; MULTICOPPER_OXIDASE2; 1.
KW Oxidoreductase.
SQ SEQUENCE 570 AA; 63121 MW; 55FC0D979D7A96AF CRC64;

Query Match 47.8%; Score 43; DB 10; Length 570;
Best Local Similarity 46.7%; Pred. No. 76;
Matches 7; Conservative 4; Mismatches 4; Indels 0; Gaps 0;

QY 2 GYGRDDNDRPPSRK 16
DB 486 GFGNFDPDKDPKFN 500

RESULT 40
Q9A6Y0
ID Q9A6Y0 PRELIMINARY; PRT; 686 AA.
AC Q9A6Y0;
DT 01-JUN-2001 (TREMblrel. 17, Created)
DT 01-JUN-2001 (TREMblrel. 17, Last sequence update)
DT 01-JUN-2002 (TREMblrel. 21, Last annotation update)
DE NADH dehydrogenase I, G subunit.
GN CC1946.
OS Caulobacter crescentus.
OC Bacteria; Proteobacteria; alpha subdivision; Caulobacter group;
OC Caulobacter.
OX NCBI_TaxID=155892;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 19089 / CB15; PubMed=11259647;
RX MEDLINE=21173698;
RA Niernan W.C., Feldblum T.V., Laub M.T., Paulsen I.T., Nelson K.E.,
  Eisen J., Heidelberg J.F., Alley M.R.K., Onta N., Maddock J.R.,
  Potocka I., Nelson W.C., Newton A., Stephens C., Phadke N.D., Ely B.,
  DeBoy R.T., Dodson R.J., Durkin A.S., Gwinn M.L., Haft D.H.,
  Kolonay J.F., Smit J., Craven M.B., Khouri H., Shetty J., Berry K.,
  Ueberback T., Tran K., Wolf A., Vamathevan J., Ermolaeva M., White O.,
  Salzberg S.L., Venter J.C., Shapiro L., Fraser C.M.;
  "Complete genome sequence of Caulobacter crescentus.";
  Proc. Natl. Acad. Sci. U.S.A. 98:4136-4141(2001).
RL EMBL; AE005868; AAK23921.1; -.
DR TIGR; CC1946; -.
DR InterPro; IPR000283; Complex1_75K.
DR InterPro; IPR001041; Ferredoxin.
DR InterPro; IPR001467; Prok_Mboxred.
DR Pfam; PF00384; molybdopterin; 1.
  
```

```

DR PROSITE; PS00641; COMPLEX1_75K_1; 1.
DR PROSITE; PS00642; COMPLEX1_75K_2; 1.
DR PROSITE; PS00643; COMPLEX1_75K_3; 1.
KW Iron-sulfur; Complete proteome.
SQ SEQUENCE 686 AA; 73163 MW; 3B4955F60FB55F70 CRC64;

Query Match 47.8%; Score 43; DB 16; Length 686;
Best Local Similarity 46.7%; Pred. No. 94;
Matches 7; Conservative 5; Mismatches 3; Indels 0; Gaps 0;

QY 1 VGYGRDDNDRPPSRK 15
DB 115 VGYGRDDSRYEENK 129
  
```

Search completed: Apr 11 9, 2003, 12:58:02  
 Job time : 17.1639 secs

**THIS PAGE BLANK (USPTO)**



GenCore version 5.1.4 p5.4578  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 9, 2003, 12:53:47 ; Search time 4.98361 Seconds  
(without alignments)  
94.463 Million cell updates/sec

Title: US-10-020-441-3  
Perfect score: 90  
Sequence: 1 VGYGRDNDPRPSRKN 16

Scoring table: BLOSUM62  
Gapop 10.0, Gapext 0.5

Searched: 262574 seqs, 29422922 residues

Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued Patents AA: \*  
1: /cgn2\_6/ptodata/1/1aa/5A COMB.pep: \*  
2: /cgn2\_6/ptodata/1/1aa/5B COMB.pep: \*  
3: /cgn2\_6/ptodata/1/1aa/6A COMB.pep: \*  
4: /cgn2\_6/ptodata/1/1aa/6B COMB.pep: \*  
5: /cgn2\_6/ptodata/1/1aa/PTUS COMB.pep: \*  
6: /cgn2\_6/ptodata/1/1aa/backfile1.pep: \*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	44	48.9	614	4 US-09-832-498-2	Sequence 2, Appli
2	44	48.9	614	4 US-09-832-614A-2	Sequence 2, Appli
3	43	47.8	258	4 US-09-134-001C-4764	Sequence 4764, Ap
4	41	45.6	168	4 US-09-512-342-20	Sequence 20, Appli
5	41	45.6	514	4 US-08-796-899-25	Sequence 25, Appli
6	41	45.6	572	2 US-08-669-524-8	Sequence 8, Appli
7	40	44.4	48	1 US-08-457-245-21	Sequence 21, Appli
8	40	44.4	228	4 US-07-724-864-46	Sequence 46, Appli
9	40	44.4	351	2 US-08-701-191A-3	Sequence 3, Appli
10	40	44.4	783	6 5231168-2	Patent No. 5231168
11	39	43.3	104	2 US-08-801-972-5	Sequence 5, Appli
12	39	43.3	104	2 US-09-178-881-5	Sequence 5, Appli
13	38	42.2	17	3 US-09-136-421-10	Sequence 10, Appli
14	38	42.2	445	4 US-08-845-258-38	Sequence 38, Appli
15	38	42.2	445	4 US-08-990-571-38	Sequence 38, Appli
16	38	42.2	445	4 US-08-723-142A-38	Sequence 38, Appli
17	38	42.2	445	4 US-09-528-784A-38	Sequence 38, Appli
18	38	42.2	462	1 US-08-267-092A-1	Sequence 1, Appli
19	38	42.2	462	2 US-08-540-412-1	Sequence 1, Appli
20	38	42.2	462	4 US-09-051-342-1	Sequence 1, Appli
21	38	42.2	462	4 US-08-468-161-1	Sequence 1, Appli
22	38	42.2	462	4 US-09-051-759-1	Sequence 1, Appli
23	38	42.2	462	5 PCT-US95-08156-1	Sequence 1, Appli
24	38	42.2	523	4 US-09-302-620B-95	Sequence 95, Appli
25	38	42.2	666	4 US-09-528-784A-85	Sequence 85, Appli
26	38	42.2	1132	4 US-09-528-784A-87	Sequence 87, Appli
27	37	41.1	43	1 US-08-098-005-1	Sequence 1, Appli

28	37	41.1	104	4 US-09-227-357-172	Sequence 172, App
29	37	41.1	207	4 US-09-068-960-43	Sequence 43, Appli
30	37	41.1	341	2 US-08-836-854-14	Sequence 14, Appli
31	37	41.1	570	2 US-08-669-524-3	Sequence 3, Appli
32	37	41.1	570	2 US-08-669-524-4	Sequence 4, Appli
33	37	41.1	570	4 US-09-068-960-2	Sequence 2, Appli
34	37	41.1	570	4 US-09-068-960-4	Sequence 4, Appli
35	37	41.1	570	4 US-09-068-960-6	Sequence 6, Appli
36	37	41.1	570	4 US-09-068-960-8	Sequence 8, Appli
37	37	41.1	570	4 US-09-068-960-10	Sequence 10, Appli
38	37	41.1	570	4 US-09-068-960-13	Sequence 13, Appli
39	37	41.1	663	1 US-08-765-081-7	Sequence 7, Appli
40	37	41.1	663	3 US-09-098-082-7	Sequence 7, Appli
41	37	41.1	694	2 US-08-701-240-2	Sequence 2, Appli
42	37	41.1	694	4 US-09-138-236-2	Sequence 2, Appli
43	37	41.1	703	5 PCT-US95-06994-9	Sequence 9, Appli
44	37	41.1	778	5 PCT-US93-03076-3	Sequence 3, Appli
45	37	41.1	778	5 PCT-US93-03076-4	Sequence 4, Appli

# ALIGNMENTS

RESULT 1  
US-09-832-498-2  
; Sequence 2, Application US/09832498  
; Patent No. 6410291  
; GENERAL INFORMATION:  
; APPLICANT: Danielson, Steffen  
; APPLICANT: Schneider, Palle  
; TITLE OF INVENTION: Polypeptides having haloperoxidase activity  
; FILE REFERENCE: 10040.200-US  
; CURRENT APPLICATION NUMBER: US/09/832.498  
; NUMBER OF SEQ ID NOS: 2  
; SOFTWARE: PatentIn version 3.0  
; SEQ ID NO 2  
; LENGTH: 614  
; TYPE: PRT  
; ORGANISM: *Geniculosporium* sp.  
US-09-832-498-2

Query Match 48.9%; Score 44; DB 4; Length 614;  
Best Local Similarity 80.0%; Pred. No. 17;  
Matches 8; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Oy 5 RDDNDPRSR 14  
||| |||:  
Db 366 RDDNFRDPR 375

RESULT 2  
US-09-832-614A-2  
; Sequence 2, Application US/09832614A  
; Patent No. 6410292  
; GENERAL INFORMATION:  
; APPLICANT: No. 6410292ozymes A/S  
; APPLICANT: Danielson, Steffen  
; APPLICANT: Schneider, Palle  
; TITLE OF INVENTION: Nucleic acids encoding polypeptides having haloperoxidase activi  
; FILE REFERENCE: 10173.200-US  
; CURRENT APPLICATION NUMBER: US/09/832.614A  
; CURRENT FILING DATE: 2001-04-11  
; NUMBER OF SEQ ID NOS: 2  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 2  
; LENGTH: 614  
; TYPE: PRT  
; ORGANISM: *Geniculosporium* sp.  
US-09-832-614A-2

Query Match 48.9%; Score 44; DB 4; Length 614;  
Best Local Similarity 80.0%; Pred. No. 17;

Matches 8; Conservative 1; Mismatches 1; Indels 0; Gaps 0;  
Qy 5 RDDNDRDPDR 14  
||| |||:  
Db 366 RDDNRPDR 375

RESULT 3  
US-09-134-001C-4764  
; Sequence 4764, Application US/09134001C  
; Patent No. 6380370  
; GENERAL INFORMATION:  
; APPLICANT: Lynn Doucette-Stamm et al  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS  
; FILE REFERENCE: GTC-007  
; CURRENT APPLICATION NUMBER: US/09/134,001C  
; PRIOR FILING DATE: 1998-08-13  
; PRIOR APPLICATION NUMBER: US 60/064,964  
; PRIOR FILING DATE: 1997-11-08  
; PRIOR APPLICATION NUMBER: US 60/055,779  
; PRIOR FILING DATE: 1997-08-14  
; NUMBER OF SEQ ID NOS: 5674  
; SEQ ID NO 4764  
; LENGTH: 258  
; TYPE: PRT  
; ORGANISM: Staphylococcus epidermidis  
US-09-134-001C-4764

Query Match 47.8%; Score 43; DB 4; Length 258;  
Best Local Similarity 77.8%; Pred. No. 9.8;  
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 2 GYGRRDND 10  
||| :|||  
Db 183 GYGRRDND 191

RESULT 4  
US-09-512-342-20  
; Sequence 20, Application US/09512342  
; Patent No. 6388068  
; GENERAL INFORMATION:  
; APPLICANT: SATOH, SHINOBU  
; APPLICANT: MASUDA, SUSUMU  
; TITLE OF INVENTION: METHOD FOR PRODUCING FOREIGN POLYPEPTIDE IN PLANT  
; FILE REFERENCE: 081356/0142  
; CURRENT APPLICATION NUMBER: US/09/512,342  
; CURRENT FILING DATE: 2000-02-24  
; NUMBER OF SEQ ID NOS: 38  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 20  
; LENGTH: 168  
; TYPE: PRT  
; ORGANISM: Cucumis sativus  
US-09-512-342-20

Query Match 45.6%; Score 41; DB 4; Length 168;  
Best Local Similarity 70.0%; Pred. No. 13;  
Matches 7; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Qy 1 VGYGRDND 10  
||| |||:  
Db 53 VGYGRDND 62

RESULT 5  
US-08-796-899-25  
; Sequence 25, Application US/08796899  
; Patent No. 6160202  
; GENERAL INFORMATION:  
; APPLICANT: BUSTOS, Mauricio M

APPLICANT: CHERN, Maw-Sheng  
; TITLE OF INVENTION: MODIFICATION OF SEED CROPS WITH  
; FILE REFERENCE: TRANSCRIPTION FACTORS  
; NUMBER OF SEQUENCES: 32  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Burns, Doane, Swecker & Mathis  
; STREET: P.O. Box 1404  
; CITY: Alexandria  
; STATE: Virginia  
; COUNTRY: United States  
; ZIP: 22313-1404

COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30

CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/796,899  
; FILING DATE: 06-FEB-1997  
; CLASSIFICATION: 800

PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/319,544  
; FILING DATE: 07-OCT-1994  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Neuth, Donna M

REGISTRATION NUMBER: 36,607  
; REFERENCE/DOCKET NUMBER: 028754-005  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (703) 836-6620  
; TELEFAX: (703) 836-2021

INFORMATION FOR SEQ ID NO: 25:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 514 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear

MOLECULE TYPE: DNA (genomic)  
US-08-796-899-25

Query Match 45.6%; Score 41; DB 4; Length 514;  
Best Local Similarity 57.1%; Pred. No. 44;  
Matches 8; Conservative 1; Mismatches 5; Indels 0; Gaps 0;

Qy 2 GYGRRDND 15  
||| |||:  
Db 271 GYGRRDND 284

RESULT 6  
US-08-669-524-8  
; Sequence 8, Application US/08669524  
; Patent No. 5843758  
; GENERAL INFORMATION:  
; APPLICANT: RUSSELL, Robyn J.  
; APPLICANT: NEWCOMB, Richard D.  
; APPLICANT: ROBIN, Geoffrey C.  
; APPLICANT: BOYCE, Thomas M.  
; APPLICANT: CAMPBELL, Peter M.  
; APPLICANT: PARKER, Anthony G.  
; APPLICANT: OAKSHOTT, John G.  
; APPLICANT: SMYTH, Kertie A.  
; TITLE OF INVENTION: ENZYME BASED BIOREMEDIATION  
; NUMBER OF SEQUENCES: 22  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Lowe Price Leblanc & Becker  
; STREET: 99 Canal Center Plaza, Suite 300  
; CITY: Alexandria  
; STATE: Virginia  
; COUNTRY: USA  
; ZIP: 22314

COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/669,524  
FILING DATE:  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Price, Robert L.  
REGISTRATION NUMBER: 22,685  
REFERENCE/DOCKET NUMBER: 1451-021  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 703-684-1111  
TELEFAX: 703-684-1124  
INFORMATION FOR SEQ ID NO: 8:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 572 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
US-08-669-524-8

Query Match 45.6%; Score 41; DB 2; Length 572;  
Best Local Similarity 63.6%; Pred. No. 49;  
Matches 7; Conservative 2; Mismatches 2; Indels 0; Gaps 0;  
QY 1 VGYGRDNDNRD 11  
Db 268 VGYGRDNDNRD 278

RESULT 7  
US-08-457-245-21  
Sequence 21, Application US/08457245  
Patent No. 5573915  
GENERAL INFORMATION:  
APPLICANT: BARRY III, Clifton E.  
TITLE OF INVENTION: CLONING AND EXPRESSION OF DNA INVOLVED  
IN THE BIOSYNTHESIS OF CYCLOPORANATED MYCOLIC ACIDS IN  
TITLE OF INVENTION: MYCOBACTERIUM TUBERCULOSIS  
NUMBER OF SEQUENCES: 21  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Townsend and Townsend Kourie and Crew  
STREET: Stewart Street Tower, One Market Plaza  
CITY: San Francisco  
STATE: California  
COUNTRY: US  
ZIP: 94105-1493  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/457,245  
FILING DATE:  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Chambers, Guy W.  
REGISTRATION NUMBER: 30,617  
REFERENCE/DOCKET NUMBER: 15280-216000  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (415) 543-9600  
TELEFAX: (415) 543-5043  
INFORMATION FOR SEQ ID NO: 21:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 48 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-457-245-21

Query Match 44.4%; Score 40; DB 1; Length 48;  
Best Local Similarity 54.5%; Pred. No. 5;  
Matches 6; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

QY 6 DDNRDPSRKN 16  
Db 29 DDDKDPSSRS 39

RESULT 8  
US-09-724-864-46  
Sequence 46, Application US/09724864  
Patent No. 6380362  
GENERAL INFORMATION:  
APPLICANT: Watson, James D.  
TITLE OF INVENTION: Polynucleotides, polypeptides expressed  
by the polynucleotides and methods for their use.  
FILE REFERENCE: 11000,105001  
CURRENT APPLICATION NUMBER: US/09/724,864  
CURRENT FILING DATE: 2000-11-28  
PRIOR APPLICATION NUMBER: U.S. No. 6380362 60/171,678  
PRIOR FILING DATE: 1999-12-23  
NUMBER OF SEQ ID NOS: 72  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 46  
LENGTH: 228  
TYPE: PRT  
ORGANISM: Mouse  
US-09-724-864-46

Query Match 44.4%; Score 40; DB 4; Length 228;  
Best Local Similarity 40.0%; Pred. No. 27;  
Matches 6; Conservative 4; Mismatches 5; Indels 0; Gaps 0;

QY 2 GYGRDNDNRDPSRKN 16  
Db 20 GMAQEDGDADPEPN 34

RESULT 9  
US-08-701-191A-3  
Sequence 3, Application US/08701191A  
Patent No. 5942428  
GENERAL INFORMATION:  
APPLICANT: Moosa Mohammadi, Joseph Schlessinger,  
and Stevan R. Hubbard  
TITLE OF INVENTION: CRYSTALS OF THE TYROSINE KINASE DOMAIN  
OF NON-INSULIN RECEPTOR TYROSINE KINASE  
NUMBER OF SEQUENCES: 41  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Lyon & Lyon  
STREET: 633 West Fifth Street  
CITY: Los Angeles  
STATE: California  
COUNTRY: U.S.A.  
ZIP: 90071-2066  
COMPUTER READABLE FORM:  
MEDIUM TYPE: 3.5" Diskette, 1.44 MB  
MEDIUM TYPE: storage  
COMPUTER: IBM compatible  
OPERATING SYSTEM: IBM P.C. DOS 5.0  
SOFTWARE: FastSeq for Windows 2.0  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/701,191A  
FILING DATE: August 21, 1996  
CLASSIFICATION: 530  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER:  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: Warburg, Richard J.

```

; REGISTRATION NUMBER: 32,327
; REFERENCE/DOCKET NUMBER: 227/088
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (213) 489-1600
; TELEFAX: (213) 955-0440
; TELEX: 67-3510
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 351 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHEICAL: NO
; US-08-701-191A-3

```

```

Query Match 44.4%; Score 40; DB 2; Length 351;
Best Local Similarity 54.5%; Pred. No. 43;
Matches 6; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

```

```

Qy 6 DDNDPSPSKN 16
Db 27 DDDKDPSSRS 37

```

```

RESULT 10
5231168-2
; Patent No. 5231168
; APPLICANT: DIEGIEL, MORTEN; BORRE, MARTIN; JEPSN, SOREN;
; VUUST, JENS; RIENECK, KLAUS; WIND, ANNETTE; JAKOBSEN, PALLE H.
; TITLE OF INVENTION: MALARIA ANTIGEN
; NUMBER OF SEQUENCES: 19
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/409,658
; FILING DATE: 18-SEP-1989
; SEQ ID NO: 2:
; LENGTH: 783
; 5231168-2

```

```

Query Match 44.4%; Score 40; DB 6; Length 783;
Best Local Similarity 46.7%; Pred. No. 1e+02;
Matches 7; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

```

```

Qy 2 GYGRDDNDPSPSKN 16
Db 715 GLSRKDEKDSNKN 729

```

```

RESULT 11
US-08-801-972-5
; Sequence 5, Application US/08801972
; Patent No. 5831018
; GENERAL INFORMATION:
; APPLICANT: Goli, Surya K.
; TITLE OF INVENTION: A NOVEL HUMAN CYTOCHROME
; NUMBER OF SEQUENCES: 5
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Incyte Pharmaceuticals, Inc.
; STREET: 3174 Porter Drive
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/801,972
; FILING DATE: Filed Herewith
; PRIOR APPLICATION DATA:

```

```

; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Billings, Lucy J.
; REGISTRATION NUMBER: 36,749
; REFERENCE/DOCKET NUMBER: PF-0216 US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-855-0555
; TELEFAX: 415-845-4166
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 104 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; LIBRARY: Genbank
; CLONE: 1217655
; US-08-801-972-5

```

```

Query Match 43.3%; Score 39; DB 2; Length 104;
Best Local Similarity 60.0%; Pred. No. 17;
Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

```

```

Qy 7 DNDPSPSKN 16
Db 62 DDDKDPSSKN 71

```

```

RESULT 12
US-09-178-881-5
; Sequence 5, Application US/09178881
; Patent No. 6083704
; GENERAL INFORMATION:
; APPLICANT: Goli, Surya K.
; TITLE OF INVENTION: A NOVEL HUMAN CYTOCHROME B5
; NUMBER OF SEQUENCES: 5
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Incyte Pharmaceuticals, Inc.
; STREET: 3174 Porter Drive
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/178,881
; FILING DATE:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/801,972
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Billings, Lucy J.
; REGISTRATION NUMBER: 36,749
; REFERENCE/DOCKET NUMBER: PF-0216 US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-855-0555
; TELEFAX: 415-845-4166
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 104 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; LIBRARY: Genbank
; CLONE: 1217655
; US-09-178-881-5

```

Query Match 43.3%; Score 39; DB 3; Length 104;  
Best Local Similarity 60.0%; Pred. No. 17;  
Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 7 DNDNDPSRKN 16  
:|:|:|:  
Db 62 DGDNDPSRKN 71

RESULT 13  
US-09-136-421-10  
Sequence 10, Application US/09136421

Patent No. 6037150  
GENERAL INFORMATION:  
APPLICANT: IATROU, Kostas  
APPLICANT: PARELLI, Patrick J.  
APPLICANT: BEHIE, Leo A.  
TITLE OF INVENTION: INSECT SEQUENCES FOR IMPROVING THE EFFICIENCY OF  
FILE REFERENCE: 028722-169  
CURRENT APPLICATION NUMBER: US/09/136,421  
EARLIER APPLICATION NUMBER: 60/056,871  
EARLIER FILING DATE: 1997-08-21  
NUMBER OF SEQ ID NOS: 11  
SOFTWARE: PatentIn Ver. 2.0  
SEQ ID NO 10  
LENGTH: 17  
TYPE: PRT  
ORGANISM: Unknown  
FEATURE:  
OTHER INFORMATION: Description of Unknown Organism: synthetic  
US-09-136-421-10

Query Match 42.2%; Score 38; DB 3; Length 17;  
Best Local Similarity 66.7%; Pred. No. 3.5;  
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 4 GRDNDNDP 12  
:|:|:|:  
Db 9 GGDNDNDP 17

RESULT 14  
US-08-845-258-38  
Sequence 38, Application US/08845258

Patent No. 6183976  
GENERAL INFORMATION:  
APPLICANT: Reed, Steven G.  
APPLICANT: Lodes, Michael J.  
APPLICANT: Houghton, Raymond  
APPLICANT: Sleeth, Paul R.  
TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THE DIAGNOSIS  
TITLE OF INVENTION: AND TREATMENT OF B. MICROTI INFECTION  
NUMBER OF SEQUENCES: 53  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: SEED AND BERRY  
STREET: 6300 Columbia Center, 701 Fifth Avenue  
CITY: Seattle  
STATE: Washington  
COUNTRY: USA  
ZIP: 98104  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/845,258  
FILING DATE: 24-APR-1997  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Maki, David J.

REGISTRATION NUMBER: 31,392  
REFERENCE/DOCKET NUMBER: 210121.426C1  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (206) 622-4900  
TELEFAX: (206) 682-6031  
INFORMATION FOR SEQ ID NO: 38:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 445 amino acids  
TYPE: amino acid  
STRANDEDNESS:  
TOPOLOGY: linear  
US-08-845-258-38

Query Match 42.2%; Score 38; DB 4; Length 445;  
Best Local Similarity 75.0%; Pred. No. 1.2e+02;  
Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 6 DNDNDPS 13  
:|:|:|:  
Db 381 EDNERDPS 388

RESULT 15  
US-08-990-571-38  
Sequence 38, Application US/08990571

Patent No. 6214971  
GENERAL INFORMATION:  
APPLICANT: Reed, Steven G. et al.  
TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THE DIAGNOSIS AND TREATMENT OF B. M  
NUMBER OF SEQUENCES: 79  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: SEED AND BERRY  
STREET: 6300 Columbia Center, 701 Fifth Avenue  
CITY: Seattle  
STATE: Washington  
COUNTRY: USA  
ZIP: 98104  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/990,571  
FILING DATE: 11-DEC-1997  
CLASSIFICATION:  
ATTORNEY/AGENT INFORMATION:  
NAME: Maki, David J.  
REGISTRATION NUMBER: 31,392  
REFERENCE/DOCKET NUMBER: 210121.426C2  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (206) 622-4900  
TELEFAX: (206) 682-6031  
INFORMATION FOR SEQ ID NO: 38:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 445 amino acids  
TYPE: amino acid  
STRANDEDNESS:  
TOPOLOGY: linear  
US-08-990-571-38

Query Match 42.2%; Score 38; DB 4; Length 445;  
Best Local Similarity 75.0%; Pred. No. 1.2e+02;  
Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 6 DNDNDPS 13  
:|:|:|:  
Db 381 EDNERDPS 388

RESULT 16  
US-08-723-142A-38  
Sequence 38, Application US/08723142A

; Patent No. 6306396  
; GENERAL INFORMATION:  
; APPLICANT: Reed, Steven G.  
; APPLICANT: Lodes, Michael J.  
; APPLICANT: Houghton, Raymond  
; APPLICANT: Sleath, Paul R.  
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THE DIAGNOSIS  
; TITLE OF INVENTION: AND TREATMENT OF B. MICROTI INFECTION  
; NUMBER OF SEQUENCES: 49  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: SEED AND BERRY  
; STREET: 6300 Columbia Center, 701 Fifth Avenue  
; CITY: Seattle  
; STATE: Washington  
; COUNTRY: USA  
; ZIP: 98104  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentin Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/723,142A  
; FILING DATE: 01-OCT-1996  
; CLASSIFICATION: 536  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Makl, David J.  
; REGISTRATION NUMBER: 31,392  
; REFERENCE/DOCKET NUMBER: 210121.426  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (206) 622-4900  
; TELEFAX: (206) 622-6031  
; INFORMATION FOR SEQ ID NO: 38:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 445 amino acids  
; TYPE: amino acid  
; STRANDEDNESS:  
; TOPOLOGY: linear  
; US-08-723-142A-38

Query Match 42.2%; Score 38; DB 4; Length 445;  
Best Local Similarity 75.0%; Pred. No. 1.2e+02;  
Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Oy 6 DDNDRDPS 13  
:|||||  
Db 381 EDNERDPS 388

RESULT 17  
US-09-528-784A-38  
; Sequence 38, Application US/09528784A  
; Patent No. 6451315  
; GENERAL INFORMATION:  
; APPLICANT: Reed, Steven G.  
; APPLICANT: Lodes, Michael J.  
; APPLICANT: Houghton, Raymond L.  
; APPLICANT: Sleath, Paul R.  
; APPLICANT: McNeill, Patricia D.  
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THE DIAGNOSIS  
; TITLE OF INVENTION: AND TREATMENT OF B. MICROTI INFECTION  
; FILE REFERENCE: 210121.426C4  
; CURRENT APPLICATION NUMBER: US/09/528,784A  
; CURRENT FILING DATE: 2000-03-17  
; NUMBER OF SEQ ID NOS: 90  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 38  
; LENGTH: 445  
; TYPE: PRT  
; ORGANISM: Babesia microti  
US-09-528-784A-38

Query Match 42.2%; Score 38; DB 4; Length 445;

Best Local Similarity 75.0%; Pred. No. 1.2e+02;  
Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;  
Oy 6 DDNDRDPS 13  
:|||||  
Db 381 EDNERDPS 388

RESULT 18  
US-08-267-092A-1  
; Sequence 1, Application US/08267092A  
; Patent No. 5596686  
; GENERAL INFORMATION:  
; APPLICANT: Defeo-Jones, Deborah  
; APPLICANT: Garsky, Victor M.  
; APPLICANT: Jones, Raymond E.  
; APPLICANT: Oliff, Allen I.  
; TITLE OF INVENTION: NOVEL PEPTIDES  
; NUMBER OF SEQUENCES: 65  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: David A. Muthard  
; STREET: 126 E. Lincoln Avenue, P.O. Box 2000  
; CITY: Rahway  
; STATE: New Jersey  
; COUNTRY: USA  
; ZIP: 07065-0900  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentin Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/267,092A  
; FILING DATE:  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Muthard, David A.  
; REGISTRATION NUMBER: 35,297  
; REFERENCE/DOCKET NUMBER: 19253  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (908) 594-3903  
; TELEFAX: (908) 594-4720  
; INFORMATION FOR SEQ ID NO: 1:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 462 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: peptide  
; HYPOTHETICAL: NO  
; ANTI-SENSE: NO  
; FRAGMENT TYPE: internal  
US-08-267-092A-1

Query Match 42.2%; Score 38; DB 1; Length 462;  
Best Local Similarity 57.1%; Pred. No. 1.2e+02;  
Matches 8; Conservative 1; Mismatches 5; Indels 0; Gaps 0;

Oy 3 YGRDNDPRDPSRKN 16  
:|||||  
Db 69 YHVDANDHDQSRKS 82

RESULT 19  
US-08-540-412-1  
; Sequence 1, Application US/08540412  
; Patent No. 5866679  
; GENERAL INFORMATION:  
; APPLICANT: Defeo-Jones, Deborah  
; APPLICANT: Feng, Dong-Mei  
; APPLICANT: Garsky, Victor M.  
; APPLICANT: Jones, Raymond E.  
; APPLICANT: Oliff, Allen I.

```

; TITLE OF INVENTION: NOVEL PEPTIDES
; NUMBER OF SEQUENCES: 194
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: DAVID A. MUTHARD
; STREET: 126 E. Lincoln Avenue, P.O. BOX 2000
; CITY: RAHWAY
; STATE: NEW JERSEY
; ZIP: 07065
; COUNTRY: U.S.A.
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/540,412
; FILING DATE:
; CLASSIFICATION: 530
; ATTORNEY/AGENT INFORMATION:
; NAME: Muthard, David A.
; REGISTRATION NUMBER: 35,297
; REFERENCE/DOCKET NUMBER: 192531C
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (908)594-3903
; TELEFAX: (908)594-4720
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 462 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; FRAGMENT TYPE: internal
; US-08-540-412-1

Query Match 42.2%; Score 38; DB 2; Length 462;
Best Local Similarity 57.1%; Pred. No. 1.2e+02;
Matches 8; Conservative 1; Mismatches 5; Indels 0; Gaps 0;

QY 3 YGRDNDRDPGRKN 16
Db 69 YHVDANDHDSGRKS 82

RESULT 20
US-09-051-342-1
; Sequence 1, Application US/09051342
; Patent No. 6130204
; GENERAL INFORMATION:
; APPLICANT: Defeo-Jones, Deborah
; APPLICANT: Feng, Dong-Mei
; APPLICANT: Garsky, Victor M.
; APPLICANT: Jones, Raymond E.
; APPLICANT: Oliff, Allen I.
; TITLE OF INVENTION: NOVEL PEPTIDES
; NUMBER OF SEQUENCES: 194
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: DAVID A. MUTHARD
; STREET: 126 E. Lincoln Avenue, P.O. BOX 2000
; CITY: RAHWAY
; STATE: NEW JERSEY
; ZIP: 07065
; COUNTRY: U.S.A.
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/051,342
; FILING DATE:

```

```

; CLASSIFICATION: 530
; ATTORNEY/AGENT INFORMATION:
; NAME: Muthard, David A.
; REGISTRATION NUMBER: 35,297
; REFERENCE/DOCKET NUMBER: 192531C
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (908)594-3903
; TELEFAX: (908)594-4720
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 462 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; FRAGMENT TYPE: internal
; US-09-051-342-1

Query Match 42.2%; Score 38; DB 4; Length 462;
Best Local Similarity 57.1%; Pred. No. 1.2e+02;
Matches 8; Conservative 1; Mismatches 5; Indels 0; Gaps 0;

QY 3 YGRDNDRDPGRKN 16
Db 69 YHVDANDHDSGRKS 82

RESULT 21
US-08-468-161-1
; Sequence 1, Application US/08468161
; Patent No. 6143864
; GENERAL INFORMATION:
; APPLICANT: Defeo-Jones, Deborah
; APPLICANT: Feng, Dong-Mei
; APPLICANT: Garsky, Victor M.
; APPLICANT: Jones, Raymond E.
; APPLICANT: Oliff, Allen I.
; TITLE OF INVENTION: NOVEL PEPTIDES
; NUMBER OF SEQUENCES: 146
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: DAVID A. MUTHARD
; STREET: 126 E. Lincoln Avenue, P.O. BOX 2000
; CITY: RAHWAY
; STATE: NEW JERSEY
; ZIP: 07065
; COUNTRY: U.S.A.
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/468,161
; FILING DATE:
; CLASSIFICATION: 530
; ATTORNEY/AGENT INFORMATION:
; NAME: Muthard, David A.
; REGISTRATION NUMBER: 35,297
; REFERENCE/DOCKET NUMBER: 192531B
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (908)594-3903
; TELEFAX: (908)594-4720
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 462 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; HYPOTHETICAL: NO
; ANTI-SENSE: NO

```

```
; FRAGMENT TYPE: internal
US-08-468-161-1

Query Match          42.2%; Score 38; DB 4; Length 462;
Best Local Similarity 57.1%; Pred. No. 1.2e+02;
Matches 8; Conservative 1; Mismatches 5; Indels 0; Gaps 0;

Qy 3 YGRDNDRDRPSRKN 16
Db 69 YHVDANDHDQSRKS 82

RESULT 22
US-09-051-759-1
; Sequence 1, Application US/09051759
; Patent No. 6177404
; GENERAL INFORMATION:
; APPLICANT: Merck & Co., Inc.
; APPLICANT: Defeo-Jones, Deborah
; APPLICANT: Garsky, Victor M.
; APPLICANT: Jones, Raymond E.
; APPLICANT: Oloff, Allen I.
; APPLICANT: Scolnick, Edward M.
; TITLE OF INVENTION: CONJUGATES USEFUL IN THE TREATMENT OF
; TITLE OF INVENTION: BENIGN PROSTATIC HYPERPLASIA
; FILE REFERENCE: 19560P
; CURRENT APPLICATION NUMBER: US/09/051,759
; PRIOR FILING DATE: 1998-08-03
; PRIOR APPLICATION NUMBER: 60/005,664
; PRIOR FILING DATE: 1995-10-18
; PRIOR APPLICATION NUMBER: PCT/US96/16490
; NUMBER OF SEQ ID NOS: 194
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1
; LENGTH: 462
; TYPE: PRT
; ORGANISM: human
US-09-051-759-1

Query Match          42.2%; Score 38; DB 4; Length 462;
Best Local Similarity 57.1%; Pred. No. 1.2e+02;
Matches 8; Conservative 1; Mismatches 5; Indels 0; Gaps 0;

Qy 3 YGRDNDRDRPSRKN 16
Db 69 YHVDANDHDQSRKS 82

RESULT 23
PCT-US95-08156-1
; Sequence 1, Application PC/TUS9508156
; GENERAL INFORMATION:
; APPLICANT: Defeo-Jones, Deborah
; APPLICANT: Feng, Dong-Mei
; APPLICANT: Garsky, Victor M.
; APPLICANT: Jones, Raymond E.
; APPLICANT: Oloff, Allen I.
; TITLE OF INVENTION: NOVEL PEPTIDES
; NUMBER OF SEQUENCES: 146
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: DAVID A. MUTTHARD
; STREET: 126 E. Lincoln Avenue, P.O. BOX 2000
; CITY: RAHWAY
; STATE: NEW JERSEY
; COUNTRY: U.S.A.
; ZIP: 07065
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: PCT/US95/08156
; FILING DATE:
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Muthard, David A.
; REGISTRATION NUMBER: 35,297
; REFERENCE/DOCKET NUMBER: 192531B
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (908)594-3903
; TELEFAX: (908)594-4720
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 462 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; HYPOTHEICAL: NO
; ANTI-SENSE: NO
; FRAGMENT TYPE: internal
PCT-US95-08156-1

Query Match          42.2%; Score 38; DB 5; Length 462;
Best Local Similarity 57.1%; Pred. No. 1.2e+02;
Matches 8; Conservative 1; Mismatches 5; Indels 0; Gaps 0;

Qy 3 YGRDNDRDRPSRKN 16
Db 69 YHVDANDHDQSRKS 82

RESULT 24
US-09-302-620B-95
; Sequence 95, Application US/09302620B
; Patent No. 6331420
; GENERAL INFORMATION:
; APPLICANT: Wilson, C. Ron
; APPLICANT: Craft, David L.
; APPLICANT: Eirich, Dudley
; APPLICANT: Eshoo, Mark
; APPLICANT: Madduri, Krishna M.
; APPLICANT: Cornett, Cathy A.
; APPLICANT: Brenner, Alfred A.
; APPLICANT: Tang, Maria
; APPLICANT: Loper, John C.
; TITLE OF INVENTION: CYTOCHROME P450 MONOOXYGENASE AND NADPH CYTOCHROME P450
; TITLE OF INVENTION: OXIDOREDUCTASE GENES AND PROTEINS RELATED TO THE OMEGA
; TITLE OF INVENTION: HYDROXYLASE COMPLEX OF CANDIDA TROPICALIS AND METHODS
; FILE REFERENCE: 1010-16.seq
; CURRENT APPLICATION NUMBER: US/09/302,620B
; CURRENT FILING DATE: 1999-04-30
; NUMBER OF SEQ ID NOS: 109
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 95
; LENGTH: 523
; TYPE: PRT
; ORGANISM: Candida tropicalis
US-09-302-620B-95

Query Match          42.2%; Score 38; DB 4; Length 523;
Best Local Similarity 58.3%; Pred. No. 1.4e+02;
Matches 7; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

Qy 3 YGRDNDRDRPSR 14
Db 437 YKXANDRPRPR 448

RESULT 25
US-09-528-784A-85
; Sequence 85, Application US/09528784A
```



Patent No. 6451315  
; GENERAL INFORMATION:  
; APPLICANT: Reed, Steven G.  
; APPLICANT: Lodes, Michael J.  
; APPLICANT: Houghton, Raymond L.  
; APPLICANT: Sleath, Paul R.  
; APPLICANT: McNeill, Patricia D.  
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THE DIAGNOSIS  
; FILE REFERENCE: 210121.426C4  
; CURRENT APPLICATION NUMBER: US/09/528,784A  
; CURRENT FILING DATE: 2000-03-17  
; NUMBER OF SEQ ID NOS: 90  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 85  
; LENGTH: 666  
; TYPE: PRT  
; ORGANISM: Babesia  
US-09-528-784A-85

Query Match 42.2%; Score 38; DB 4; Length 666;  
Best Local Similarity 75.0%; Pred. No. 1.8e+02;  
Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 6 DDNDRDPS 13  
Db 602 EDNERDPS 609

RESULT 26  
US-09-528-784A-87  
; Sequence 87, Application US/09528784A  
; Patent No. 6451315  
; GENERAL INFORMATION:  
; APPLICANT: Reed, Steven G.  
; APPLICANT: Lodes, Michael J.  
; APPLICANT: Houghton, Raymond L.  
; APPLICANT: Sleath, Paul R.  
; APPLICANT: McNeill, Patricia D.  
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THE DIAGNOSIS  
; FILE REFERENCE: 210121.426C4  
; CURRENT APPLICATION NUMBER: US/09/528,784A  
; CURRENT FILING DATE: 2000-03-17  
; NUMBER OF SEQ ID NOS: 90  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 87  
; LENGTH: 1132  
; TYPE: PRT  
; ORGANISM: Babesia  
US-09-528-784A-87

Query Match 42.2%; Score 38; DB 4; Length 1132;  
Best Local Similarity 75.0%; Pred. No. 3.2e+02;  
Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 6 DDNDRDPS 13  
Db 1068 EDNERDPS 1075

RESULT 27  
US-08-098-005-1  
; Sequence 1, Application US/08098005  
; Patent No. 5491129  
; GENERAL INFORMATION:  
; APPLICANT: SHALITEL, Shmuel  
; TITLE OF INVENTION: SYNTHETIC PEPTIDES DERIVED FROM  
; TITLE OF INVENTION: VITRONECTIN AND PHARMACEUTICAL COMPOSITIONS COMPRISING  
; NUMBER OF SEQUENCES: 1  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: BROWDY AND NEIMARK  
;

STREET: 419 Seventh Street, N.W.  
; CITY: Washington  
; STATE: D.C.  
; COUNTRY: USA  
; ZIP: 20004  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/098,005  
; FILING DATE: 29-JUL-1993  
; CLASSIFICATION: 514  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: IL 102688  
; FILING DATE: 30-JUL-1992  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: IL 104296  
; FILING DATE: 31-DEC-1992  
; ATTORNEY/AGENT INFORMATION:  
; NAME: NEIMARK, Sheridan  
; REGISTRATION NUMBER: 20,520  
; REFERENCE/DOCKET NUMBER: SHALITEL=2  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 202-628-5197  
; TELEFAX: 202-737-3528  
; TELEX: 248633  
; INFORMATION FOR SEQ ID NO: 1:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 43 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: peptide  
US-08-098-005-1

Query Match 41.1%; Score 37; DB 1; Length 43;  
Best Local Similarity 63.6%; Pred. No. 14;  
Matches 7; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 4 GRDNDRDPSPR 14  
Db 29 GRNDSRRSPR 39

RESULT 28  
US-09-227-357-172  
; Sequence 172, Application US/09227357  
; Patent No. 6342581  
; GENERAL INFORMATION:  
; APPLICANT: Fischer et al.  
; TITLE OF INVENTION: 123 Human Secreted Proteins  
; FILE REFERENCE: P2010P1  
; CURRENT APPLICATION NUMBER: US/09/227,357  
; CURRENT FILING DATE: 1999-01-08  
; EARLIER APPLICATION NUMBER: PCT/US98/13684  
; EARLIER FILING DATE: 1998-07-07  
; EARLIER APPLICATION NUMBER: 60/051,926  
; EARLIER FILING DATE: 1997-07-08  
; EARLIER APPLICATION NUMBER: 60/052,793  
; EARLIER FILING DATE: 1997-07-08  
; EARLIER APPLICATION NUMBER: 60/051,925  
; EARLIER FILING DATE: 1997-07-08  
; EARLIER APPLICATION NUMBER: 60/052,803  
; EARLIER FILING DATE: 1997-07-08  
; EARLIER APPLICATION NUMBER: 60/052,732  
; EARLIER FILING DATE: 1997-07-08  
; EARLIER APPLICATION NUMBER: 60/051,931  
; EARLIER FILING DATE: 1997-07-08  
; EARLIER APPLICATION NUMBER: 60/051,932

EARLIER FILING DATE: 1997-07-08  
EARLIER APPLICATION NUMBER: 60/051,916  
EARLIER FILING DATE: 1997-07-08  
EARLIER APPLICATION NUMBER: 60/051,930  
EARLIER FILING DATE: 1997-07-08  
EARLIER APPLICATION NUMBER: 60/051,918  
EARLIER FILING DATE: 1997-07-08  
EARLIER APPLICATION NUMBER: 60/051,920  
EARLIER FILING DATE: 1997-07-08  
EARLIER APPLICATION NUMBER: 60/052,733  
EARLIER FILING DATE: 1997-07-08  
EARLIER APPLICATION NUMBER: 60/052,795  
EARLIER FILING DATE: 1997-07-08  
EARLIER APPLICATION NUMBER: 60/051,919  
EARLIER FILING DATE: 1997-07-08  
EARLIER APPLICATION NUMBER: 60/051,928  
EARLIER FILING DATE: 1997-07-08  
EARLIER APPLICATION NUMBER: 60/055,722  
EARLIER FILING DATE: 1997-08-18  
EARLIER APPLICATION NUMBER: 60/055,723  
EARLIER FILING DATE: 1997-08-18  
EARLIER APPLICATION NUMBER: 60/055,948  
EARLIER FILING DATE: 1997-08-18  
EARLIER APPLICATION NUMBER: 60/055,949  
EARLIER FILING DATE: 1997-08-18  
EARLIER APPLICATION NUMBER: 60/055,953  
EARLIER FILING DATE: 1997-08-18  
EARLIER APPLICATION NUMBER: 60/055,950  
EARLIER FILING DATE: 1997-08-18  
EARLIER APPLICATION NUMBER: 60/055,947  
EARLIER FILING DATE: 1997-08-18  
EARLIER APPLICATION NUMBER: 60/055,964  
EARLIER FILING DATE: 1997-08-18  
EARLIER APPLICATION NUMBER: 60/056,360  
EARLIER FILING DATE: 1997-08-18  
EARLIER APPLICATION NUMBER: 60/055,684  
EARLIER FILING DATE: 1997-08-18  
EARLIER APPLICATION NUMBER: 60/055,984  
EARLIER FILING DATE: 1997-08-18  
EARLIER APPLICATION NUMBER: 60/055,954  
EARLIER FILING DATE: 1997-08-18  
EARLIER APPLICATION NUMBER: 60/058,785  
EARLIER FILING DATE: 1997-09-12  
EARLIER APPLICATION NUMBER: 60/058,664  
EARLIER FILING DATE: 1997-09-12  
EARLIER APPLICATION NUMBER: 60/058,660  
EARLIER FILING DATE: 1997-09-12  
EARLIER APPLICATION NUMBER: 60/058,661  
EARLIER FILING DATE: 1997-09-12  
NUMBER OF SEQ ID NOS: 672  
SOFTWARE: PatentIn Ver. 2.0  
SEQ ID NO 172  
LENGTH: 104  
TYPE: PRT  
ORGANISM: Homo sapiens  
US-09-227-357-172

Query Match 41.1%; Score 37; DB 4; Length 104;  
Best Local Similarity 50.0%; Pred. No. 36;  
Matches 6; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

QY 3 GYGRDDNDPSR 14  
|||:|  
Db 93 WKDDQDHVPSR 104

RESULT 29  
US-09-068-960-43  
Sequence 43, Application US/09068960A  
Patent No. 6235515  
GENERAL INFORMATION:  
APPLICANT: Commonwealth Scientific and Industrial Resrch. Org.  
TITLE OF INVENTION: MALATHION CARBOXYLSTERASE

FILE REFERENCE: Attorney Docket No. 6235515 50179-051  
CURRENT APPLICATION NUMBER: US/09/068,960A  
CURRENT FILING DATE: 1998-05-20  
EARLIER APPLICATION NUMBER: PCT/AU96/00746  
EARLIER FILING DATE: 1996-11-22  
EARLIER APPLICATION NUMBER: AU 6751  
EARLIER FILING DATE: 1995-11-23  
NUMBER OF SEQ ID NOS: 43  
SOFTWARE: PatentIn Ver. 2.0  
SEQ ID NO 43  
LENGTH: 207  
TYPE: PRT  
ORGANISM: Lucilia cuprina  
US-09-068-960-43

Query Match 41.1%; Score 37; DB 4; Length 207;  
Best Local Similarity 60.0%; Pred. No. 75;  
Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 2 GYGRDDNDRD 11  
|||:|  
Db 172 GYKGEDNDKD 181

RESULT 30  
US-08-836-854-14  
Sequence 14, Application US/08836854  
Patent No. 5824547  
GENERAL INFORMATION:  
APPLICANT: HASHINO, Kimikazu  
APPLICANT: MATSUSHITA, Hideyuki  
APPLICANT: KATO, Ikumoshin  
TITLE OF INVENTION: METHOD OF PRODUCTION OF TRANSFECTED CELLS  
NUMBER OF SEQUENCES: 21  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Browdy and Neimark  
STREET: 419 Seventh Street N.W. Ste. 300  
CITY: Washington  
STATE: D.C.  
COUNTRY: USA  
ZIP: 20004  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/836,854  
FILING DATE:  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: PCT/JP95/02425  
FILING DATE: 29-NOV-1995  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 317721/1994  
FILING DATE: 29-NOV-1994  
ATTORNEY/AGENT INFORMATION:  
NAME: Browdy, Roger L.  
REGISTRATION NUMBER: 25,618  
REFERENCE/DOCKET NUMBER: HASHINO-1  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (202) 628-5197  
TELEFAX: (202) 737-3528  
INFORMATION FOR SEQ ID NO: 14:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 341 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
US-08-836-854-14

Query Match 41.1%; Score 37; DB 2; Length 341;

Best Local Similarity 63.6%; Pred. No. 1.3e+02;  
Matches 7; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 4 GRDDNDPDR 14  
Db 314 GRNDSRRPSR 324

RESULT 31  
US-08-669-524-3

; Sequence 3, Application US/08669524

; Patent No. 5843758

; GENERAL INFORMATION:

; APPLICANT: RUSSELL, Robyn J.

; APPLICANT: NEWCOMB, Richard D.

; APPLICANT: ROBIN, Geoffrey C.

; APPLICANT: BOYCE, Thomas M.

; APPLICANT: CAMPBELL, Peter M.

; APPLICANT: PARKER, Anthony G.

; APPLICANT: OAKSHOTT, John G.

; APPLICANT: SMYTH, Kerrie A.

; TITLE OF INVENTION: ENZYME BASED BIOREMEDIATION

; NUMBER OF SEQUENCES: 22

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Lowe Price Leblanc & Becker

; STREET: 99 Canal Center Plaza, Suite 300

; CITY: Alexandria

; STATE: Virginia

; COUNTRY: USA

; ZIP: 22314

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: Patentn Release #1.0, Version #1.30

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/669,524

; FILING DATE:

; CLASSIFICATION: 435

; ATTORNEY/AGENT INFORMATION:

; NAME: Price, Robert L.

; REGISTRATION NUMBER: 22,685

; REFERENCE/DOCKET NUMBER: 1451-021

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: 703-684-1111

; TELEFAX: 703-684-1124

; INFORMATION FOR SEQ ID NO: 3:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 570 amino acids

; TYPE: amino acid

; STRANDEDNESS: single

; TOPOLOGY: linear

; MOLECULE TYPE: peptide

; US-08-669-524-3

Query Match 41.1%; Score 37; DB 2; Length 570;  
Best Local Similarity 60.0%; Pred. No. 2.3e+02;  
Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 2 GYGRDDNDRD 11  
Db 268 GYKGEDNDKD 277

RESULT 32  
US-08-669-524-4

; Sequence 4, Application US/08669524

; Patent No. 5843758

; GENERAL INFORMATION:

; APPLICANT: RUSSELL, Robyn J.

; APPLICANT: NEWCOMB, Richard D.

; APPLICANT: ROBIN, Geoffrey C.

; APPLICANT: BOYCE, Thomas M.

; APPLICANT: CAMPBELL, Peter M.

; APPLICANT: PARKER, Anthony G.

; APPLICANT: OAKSHOTT, John G.

; APPLICANT: SMYTH, Kerrie A.

; TITLE OF INVENTION: ENZYME BASED BIOREMEDIATION

; NUMBER OF SEQUENCES: 22

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Lowe Price Leblanc & Becker

; STREET: 99 Canal Center Plaza, Suite 300

; CITY: Alexandria

; STATE: Virginia

; COUNTRY: USA

; ZIP: 22314

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: Patentn Release #1.0, Version #1.30

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/669,524

; FILING DATE:

; CLASSIFICATION: 435

; ATTORNEY/AGENT INFORMATION:

; NAME: Price, Robert L.

; REGISTRATION NUMBER: 22,685

; REFERENCE/DOCKET NUMBER: 1451-021

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: 703-684-1111

; TELEFAX: 703-684-1124

; INFORMATION FOR SEQ ID NO: 4:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 570 amino acids

; TYPE: amino acid

; STRANDEDNESS: single

; TOPOLOGY: linear

; MOLECULE TYPE: peptide

Query Match 41.1%; Score 37; DB 2; Length 570;  
Best Local Similarity 60.0%; Pred. No. 2.3e+02;  
Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 2 GYGRDDNDRD 11  
Db 268 GYKGEDNDKD 277

RESULT 33  
US-09-068-960-2

; Sequence 2, Application US/09068960A

; Patent No. 623515

; GENERAL INFORMATION:

; APPLICANT: Commonwealth Scientific and Industrial Rerch. Org.

; TITLE OF INVENTION: MALATHION CARBOXYLSTERASE

; FILE REFERENCE: Attorney Docket No. 623515 50179-051

; CURRENT APPLICATION NUMBER: US/09/068,960A

; EARLIER FILING DATE: 1998-05-20

; EARLIER APPLICATION NUMBER: PCT/AU96/00746

; EARLIER FILING DATE: 1996-11-22

; EARLIER APPLICATION NUMBER: AU 6751

; NUMBER OF SEQ ID NOS: 43

; SOFTWARE: Patentn Ver. 2.0

; SEQ ID NO 2

; LENGTH: 570

; TYPE: PRT

; ORGANISM: Lucilia cuprina

Query Match 41.1%; Score 37; DB 4; Length 570;  
Best Local Similarity 60.0%; Pred. No. 2.3e+02;  
Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

US-09-068-960-2

QY 2 GYGRDNDND 11  
||:||||  
Db 268 GYKGEDNDKD 277

RESULT 34  
US-09-068-960-4  
; Sequence 4, Application US/09068960A  
; Patent No. 6235515  
; GENERAL INFORMATION:  
; APPLICANT: Commonwealth Scientific and Industrial Rerch. Org.  
; FILE REFERENCE: MALATHION CARBOXYLESTERASE  
; CURRENT APPLICATION NUMBER: US/09/068,960A  
; CURRENT FILING DATE: 1998-05-20  
; EARLIER APPLICATION NUMBER: PCT/AU96/00746  
; EARLIER FILING DATE: 1996-11-22  
; EARLIER APPLICATION NUMBER: AU 6751  
; EARLIER FILING DATE: 1995-11-23  
; NUMBER OF SEQ ID NOS: 43  
; SOFTWARE: Patentln Ver. 2.0  
; SEQ ID NO 4  
; LENGTH: 570  
; TYPE: PRT  
; ORGANISM: Lucilia cuprina  
US-09-068-960-4

Query Match 41.1%; Score 37; DB 4; Length 570;  
Best Local Similarity 60.0%; Pred. No. 2.3e+02;  
Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 2 GYGRDNDND 11  
||:||||  
Db 268 GYKGEDNDKD 277

RESULT 35  
US-09-068-960-6  
; Sequence 6, Application US/09068960A  
; Patent No. 6235515  
; GENERAL INFORMATION:  
; APPLICANT: Commonwealth Scientific and Industrial Rerch. Org.  
; TITLE OF INVENTION: MALATHION CARBOXYLESTERASE  
; FILE REFERENCE: Attorney Docket No. 6235515 50179-051  
; CURRENT APPLICATION NUMBER: US/09/068,960A  
; CURRENT FILING DATE: 1998-05-20  
; EARLIER APPLICATION NUMBER: PCT/AU96/00746  
; EARLIER FILING DATE: 1996-11-22  
; EARLIER APPLICATION NUMBER: AU 6751  
; EARLIER FILING DATE: 1995-11-23  
; NUMBER OF SEQ ID NOS: 43  
; SOFTWARE: Patentln Ver. 2.0  
; SEQ ID NO 6  
; LENGTH: 570  
; TYPE: PRT  
; ORGANISM: Lucilia cuprina  
US-09-068-960-6

Query Match 41.1%; Score 37; DB 4; Length 570;  
Best Local Similarity 60.0%; Pred. No. 2.3e+02;  
Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 2 GYGRDNDND 11  
||:||||  
Db 268 GYKGEDNDKD 277

RESULT 36  
US-09-068-960-8  
; Sequence 8, Application US/09068960A  
; Patent No. 6235515  
; GENERAL INFORMATION:  
; APPLICANT: Commonwealth Scientific and Industrial Rerch. Org.

; TITLE OF INVENTION: MALATHION CARBOXYLESTERASE  
; FILE REFERENCE: Attorney Docket No. 6235515 50179-051  
; CURRENT APPLICATION NUMBER: US/09/068,960A  
; CURRENT FILING DATE: 1998-05-20  
; EARLIER APPLICATION NUMBER: PCT/AU96/00746  
; EARLIER FILING DATE: 1996-11-22  
; EARLIER APPLICATION NUMBER: AU 6751  
; EARLIER FILING DATE: 1995-11-23  
; NUMBER OF SEQ ID NOS: 43  
; SOFTWARE: Patentln Ver. 2.0  
; SEQ ID NO 8  
; LENGTH: 570  
; TYPE: PRT  
; ORGANISM: Lucilia cuprina  
US-09-068-960-8

Query Match 41.1%; Score 37; DB 4; Length 570;  
Best Local Similarity 60.0%; Pred. No. 2.3e+02;  
Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 2 GYGRDNDND 11  
||:||||  
Db 268 GYKGEDNDKD 277

RESULT 37  
US-09-068-960-10  
; Sequence 10, Application US/09068960A  
; Patent No. 6235515  
; GENERAL INFORMATION:  
; APPLICANT: Commonwealth Scientific and Industrial Rerch. Org.  
; TITLE OF INVENTION: MALATHION CARBOXYLESTERASE  
; FILE REFERENCE: Attorney Docket No. 6235515 50179-051  
; CURRENT APPLICATION NUMBER: US/09/068,960A  
; CURRENT FILING DATE: 1998-05-20  
; EARLIER APPLICATION NUMBER: PCT/AU96/00746  
; EARLIER FILING DATE: 1996-11-22  
; EARLIER APPLICATION NUMBER: AU 6751  
; EARLIER FILING DATE: 1995-11-23  
; NUMBER OF SEQ ID NOS: 43  
; SOFTWARE: Patentln Ver. 2.0  
; SEQ ID NO 10  
; LENGTH: 570  
; TYPE: PRT  
; ORGANISM: Lucilia cuprina  
US-09-068-960-10

Query Match 41.1%; Score 37; DB 4; Length 570;  
Best Local Similarity 60.0%; Pred. No. 2.3e+02;  
Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 2 GYGRDNDND 11  
||:||||  
Db 268 GYKGEDNDKD 277

RESULT 38  
US-09-068-960-13  
; Sequence 13, Application US/09068960A  
; Patent No. 6235515  
; GENERAL INFORMATION:  
; APPLICANT: Commonwealth Scientific and Industrial Rerch. Org.  
; TITLE OF INVENTION: MALATHION CARBOXYLESTERASE  
; FILE REFERENCE: Attorney Docket No. 6235515 50179-051  
; CURRENT APPLICATION NUMBER: US/09/068,960A  
; CURRENT FILING DATE: 1998-05-20  
; EARLIER APPLICATION NUMBER: PCT/AU96/00746  
; EARLIER FILING DATE: 1996-11-22  
; EARLIER APPLICATION NUMBER: AU 6751  
; EARLIER FILING DATE: 1995-11-23  
; NUMBER OF SEQ ID NOS: 43  
; SOFTWARE: Patentln Ver. 2.0  
; SEQ ID NO 13

```
; LENGTH: 570
; TYPE: PRT
; ORGANISM: Musca domestica
US-09-068-960-13

Query Match          41.1%; Score 37; DB 4; Length 570;
Best Local Similarity 54.5%; Pred. No. 2.3e+02;
Matches 6; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

QY      1 GYGGRDNDRD 11
      ||| :||:|
Db      267 VGKGEDNEKD 277

RESULT 39
US-08-765-081-7
; Sequence 7, Application US/08765081
; Patent No. 5798260
; GENERAL INFORMATION:
; APPLICANT: Tarr, P.I., Bilge, S.S., Besser, T.E., Vary Jr., J.C.
; TITLE OF INVENTION: Escherichia Coli 0157:H7 Epithelial Adhesin
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Christensen, O'Connor, Johnson and Kindness
; STREET: 2800 Pacific First Center, 1420 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98101-2347
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette-3.5 inch, 1.44mb storage
; COMPUTER: IBM PC compatible/Pentium
; OPERATING SYSTEM: MS-Windows 3.1
; SOFTWARE: Word for Windows-6.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/765,081
; FILING DATE: March 26, 1997
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/US95/06994
; FILING DATE: June 7, 1995
; APPLICATION NUMBER: US 08/265,714
; FILING DATE: June 24, 1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Shelton, Dennis K.
; REGISTRATION NUMBER: 26,997
; TELEPHONE: 1-206-682-8100; 1-206-224-0718 (direct)
; TELEFAX: 1-206-224-0779
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 663 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; DESCRIPTION: E. coli C17A protein amino acid sequence
; HYPOTHETICAL: NO
; ORIGINAL SOURCE:
; ORGANISM: Escherichia Coli
; US-08-765-081-7

Query Match          41.1%; Score 37; DB 1; Length 663;
Best Local Similarity 70.0%; Pred. No. 2.7e+02;
Matches 7; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY      2 GYGGRDNDRD 11
      ||| | |||
Db      254 GYGFRDNRD 263

RESULT 40
US-09-098-082-7
; Sequence 7, Application US/09098082
```

```
; Patent No. 6040421
; GENERAL INFORMATION:
; APPLICANT: Tarr, P.I., Bilge, S.S., Besser, T.E., Vary Jr., J.C.
; TITLE OF INVENTION: Escherichia Coli 0157:H7 Epithelial Adhesin
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Christensen, O'Connor, Johnson and Kindness
; STREET: 2800 Pacific First Center, 1420 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98101-2347
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette-3.5 inch, 1.44mb storage
; COMPUTER: IBM PC compatible/Pentium II
; OPERATING SYSTEM: MS-Windows 95
; SOFTWARE: Word for Windows-6.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/098,082
; FILING DATE:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/765,081
; FILING DATE: March 26, 1997
; APPLICATION NUMBER: PCT/US95/06994
; FILING DATE: June 7, 1995
; APPLICATION NUMBER: US 08/265,714
; FILING DATE: June 24, 1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Sheiness, Diana K.
; REGISTRATION NUMBER: 35,356
; TELEPHONE: 1-206-682-8100; 1-206-224-0735 (direct)
; TELEFAX: 1-206-224-0779
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 663 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; DESCRIPTION: E. coli C17A protein amino acid sequence
; HYPOTHETICAL: NO
; ORIGINAL SOURCE:
; ORGANISM: Escherichia Coli
; US-09-098-082-7

Query Match          41.1%; Score 37; DB 3; Length 663;
Best Local Similarity 70.0%; Pred. No. 2.7e+02;
Matches 7; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY      2 GYGGRDNDRD 11
      ||| | |||
Db      254 GYGFRDNRD 263

Search completed: April 9, 2003, 12:59:44
Job time : 5.98361 secs
```

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.4.p5.4578  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 9, 2003, 12:53:47 ; Search time 33.0164 Seconds  
(without alignments)  
94.463 Million cell updates/sec

Title: US-10-020-441-2

Perfect score: 552  
Sequence: 1 VSFPLRLNGDQGIHQPSG.....YGRDDNDPRSRKNGILKK 106

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 262574 seqs, 29422922 residues

Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued Patents\_AA:\*  
1: /cgn2\_6/ptodata/1/1aa/5A.COMB.pep:\*  
2: /cgn2\_6/ptodata/1/1aa/5B.COMB.pep:\*  
3: /cgn2\_6/ptodata/1/1aa/6A.COMB.pep:\*  
4: /cgn2\_6/ptodata/1/1aa/6B.COMB.pep:\*  
5: /cgn2\_6/ptodata/1/1aa/PTOS.COMB.pep:\*  
6: /cgn2\_6/ptodata/1/1aa/backfilest1.pep:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	102	18.5	276	2	US-09-016-366A-15 Sequence 15, Appl
2	102	18.5	276	2	US-08-978-404B-21 Sequence 21, Appl
3	101.5	18.4	190	2	US-08-845-998-4 Sequence 4, Appl
4	101.5	18.4	190	2	US-08-845-998-6 Sequence 6, Appl
5	101.5	18.4	190	4	US-09-206-537-4 Sequence 4, Appl
6	101.5	18.4	190	4	US-09-206-537-6 Sequence 6, Appl
7	101.5	18.4	190	4	US-09-430-854-4 Sequence 4, Appl
8	101.5	18.4	190	4	US-09-430-854-6 Sequence 6, Appl
9	101	18.3	270	2	US-08-978-404B-8 Sequence 8, Appl
10	100	18.1	245	4	US-08-944-483-69 Sequence 69, Appl
11	100	18.1	275	2	US-09-016-366A-17 Sequence 17, Appl
12	100	18.1	275	2	US-08-978-404B-12 Sequence 12, Appl
13	99	17.9	273	2	US-09-016-366A-19 Sequence 19, Appl
14	99	17.9	273	2	US-08-978-404B-6 Sequence 6, Appl
15	99	17.9	273	2	US-08-978-404B-14 Sequence 14, Appl
16	95.5	17.3	242	4	US-08-944-483-58 Sequence 58, Appl
17	93	16.8	245	4	US-09-079-970A-6 Sequence 6, Appl
18	93	16.8	249	4	US-09-079-970A-5 Sequence 5, Appl
19	93	16.8	267	2	US-09-016-366A-23 Sequence 23, Appl
20	93	16.8	267	2	US-08-978-404B-18 Sequence 18, Appl
21	93	16.8	274	2	US-09-016-366A-21 Sequence 21, Appl
22	93	16.8	274	2	US-08-978-404B-16 Sequence 16, Appl
23	92.5	16.8	242	4	US-08-944-483-57 Sequence 57, Appl
24	89	16.1	274	2	US-08-978-404B-5 Sequence 5, Appl
25	85.5	15.5	269	2	US-08-978-404B-10 Sequence 10, Appl
26	83.5	15.1	290	4	US-09-386-653A-7 Sequence 7, Appl
27	83.5	15.1	315	4	US-09-386-653A-9 Sequence 9, Appl

28	83	15.0	273	2	US-08-978-404B-3 Sequence 3, Appl
29	82.5	14.9	855	2	US-09-027-337-2 Sequence 2, Appl
30	82.5	14.9	855	4	US-09-644-600-2 Sequence 12, Appl
31	81.5	14.8	224	1	US-08-278-091-12 Sequence 12, Appl
32	81.5	14.8	224	1	US-08-483-859-12 Sequence 12, Appl
33	81.5	14.8	224	1	US-08-472-173-12 Sequence 12, Appl
34	81.5	14.8	224	2	US-08-487-167-12 Sequence 12, Appl
35	81.5	14.8	224	2	US-08-483-816-12 Sequence 12, Appl
36	81.5	14.8	224	2	US-08-296-149-12 Sequence 12, Appl
37	81.5	14.8	224	2	US-08-801-499-12 Sequence 12, Appl
38	81.5	14.8	224	2	US-08-615-271-12 Sequence 12, Appl
39	81.5	14.8	224	3	US-09-074-660-12 Sequence 12, Appl
40	81.5	14.8	224	3	US-09-074-659-12 Sequence 12, Appl
41	81.5	14.8	224	3	US-09-106-468-12 Sequence 12, Appl
42	81.5	14.8	224	4	US-09-106-468A-12 Sequence 12, Appl
43	81.5	14.8	224	4	US-09-106-467-12 Sequence 12, Appl
44	81.5	14.8	247	2	US-08-978-404B-47 Sequence 47, Appl
45	80	14.5	228	2	US-08-766-982-11 Sequence 11, Appl

ALIGNMENTS

RESULT 1  
US-09-016-366A-15  
Sequence 15, Application US/09016366A  
Patent No. 5955431  
GENERAL INFORMATION:  
APPLICANT: Stevens, Richard L.  
APPLICANT: Huang, Chifu  
TITLE OF INVENTION: MAST CELL PROTEASE PEPTIDE  
TITLE OF INVENTION: INHIBITORS  
NUMBER OF SEQUENCES: 65  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Wolf, Greenfield & Sacks, P.C.  
STREET: 600 Atlantic Avenue  
CITY: Boston  
STATE: MA  
COUNTRY: U.S.A.  
ZIP: 02210-2211  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: DOS  
SOFTWARE: FastSeq for Windows Version 2.0  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/016,366A  
FILING DATE: January 30, 1998  
CLASSIFICATION: 530  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 60/037,090  
FILING DATE: 05-FEB-1997  
ATTORNEY/AGENT INFORMATION:  
NAME: Plummer, Elizabeth R.  
REGISTRATION NUMBER: 36,637  
REFERENCE/DOCKET NUMBER: B0801/7093  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 617-720-3500  
TELEFAX: 617-720-2441  
TELEX:  
INFORMATION FOR SEQ ID NO: 15:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 276 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-09-016-366A-15  
Query Match 18.5%; Score 102; DB 2; Length 276;  
Best Local Similarity 43.8%; Pred. No. 0.00017;  
Matches 21; Conservative 9; Mismatches 18; Indels 0; Gaps 0;





REFERENCE/DOCKET NUMBER: L0461/7008  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (617)720-3500  
TELEFAX: (617)720-2441  
INFORMATION FOR SEQ ID NO: 6:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 190 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-845-998-6

Query Match 18.4%; Score 101.5; DB 2; Length 190;  
Best Local Similarity 39.7%; Pred. No. 0.00012;  
Matches 25; Conservative 10; Mismatches 19; Indels 9; Gaps 2;

QY 40 IAGTSGFDIAVMAQWNLQSGIRVSLPPGSDIPPGTGVFIYGYGRDNDRDPSPRK 99  
DB 115 IIQT--GADIALLEBEPVNISSHIHTVLPASSTFPFGMPGWCWVGWDVGN----- 165

QY 100 NGG 102  
DB 166 NGG 168

## RESULT 5

US-09-206-537-4  
Sequence 4, Application US/09206537  
Patent No. 6130052  
GENERAL INFORMATION:  
APPLICANT: Van Baren, Nicolas  
APPLICANT: Coulie, Pierre G.  
APPLICANT: De Smet, Charles  
APPLICANT: Lucas, Sophie  
APPLICANT: Boon, Thierry  
TITLE OF INVENTION: LEUKEMIA ASSOCIATED GENES  
NUMBER OF SEQUENCES: 16  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Wolf, Greenfield & Sacks, P.C.  
STREET: 600 Atlantic Avenue  
CITY: Boston  
STATE: MA  
COUNTRY: US  
ZIP: 02210  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/206,537  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/845,998  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: Van Amsterdam, John R.  
REGISTRATION NUMBER: 40,212  
REFERENCE/DOCKET NUMBER: L0461/7008  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (617)720-3500  
TELEFAX: (617)720-2441  
INFORMATION FOR SEQ ID NO: 4:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 190 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-09-206-537-4

Query Match 18.4%; Score 101.5; DB 4; Length 190;  
Best Local Similarity 39.7%; Pred. No. 0.00012;

Matches 25; Conservative 10; Mismatches 19; Indels 9; Gaps 2;  
QY 40 IAGTSGFDIAVMAQWNLQSGIRVSLPPGSDIPPGTGVFIYGYGRDNDRDPSPRK 99  
DB 115 IIQT--GADIALLEBEPVNISSHIHTVLPASSTFPFGMPGWCWVGWDVGN----- 165

QY 100 NGG 102  
DB 166 NGG 168

## RESULT 6

US-09-206-537-6  
Sequence 6, Application US/09206537  
Patent No. 6130052  
GENERAL INFORMATION:  
APPLICANT: Van Baren, Nicolas  
APPLICANT: Coulie, Pierre G.  
APPLICANT: De Smet, Charles  
APPLICANT: Lucas, Sophie  
APPLICANT: Boon, Thierry  
TITLE OF INVENTION: LEUKEMIA ASSOCIATED GENES  
NUMBER OF SEQUENCES: 16  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Wolf, Greenfield & Sacks, P.C.  
STREET: 600 Atlantic Avenue  
CITY: Boston  
STATE: MA  
COUNTRY: US  
ZIP: 02210  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/206,537  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/845,998  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: Van Amsterdam, John R.  
REGISTRATION NUMBER: 40,212  
REFERENCE/DOCKET NUMBER: L0461/7008  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (617)720-3500  
TELEFAX: (617)720-2441  
INFORMATION FOR SEQ ID NO: 6:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 190 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-09-206-537-6

Query Match 18.4%; Score 101.5; DB 4; Length 190;  
Best Local Similarity 39.7%; Pred. No. 0.00012;  
Matches 25; Conservative 10; Mismatches 19; Indels 9; Gaps 2;

QY 40 IAGTSGFDIAVMAQWNLQSGIRVSLPPGSDIPPGTGVFIYGYGRDNDRDPSPRK 99  
DB 115 IIQT--GADIALLEBEPVNISSHIHTVLPASSTFPFGMPGWCWVGWDVGN----- 165

QY 100 NGG 102  
DB 166 NGG 168

## RESULT 7

US-09-430-854-4  
Sequence 4, Application US/09430854

```
/ Patent No. 6271019
/
/ GENERAL INFORMATION:
/ APPLICANT: Van Baren, Nicolas
/ APPLICANT: Coulie, Pierre G.
/ APPLICANT: De Smet, Charles
/ APPLICANT: Lucas, Sophie
/ APPLICANT: Boon, Thierry
/ TITLE OF INVENTION: LEUKEMIA ASSOCIATED GENES
/ NUMBER OF SEQUENCES: 16
/ CORRESPONDENCE ADDRESS:
/ ADDRESSER: Wolf, Greenfield & Sacks, P.C.
/ STREET: 600 Atlantic Avenue
/ CITY: Boston
/ STATE: MA
/ COUNTRY: US
/ ZIP: 02210
/
/ COMPUTER READABLE FORM:
/ MEDIUM TYPE: Floppy disk
/ COMPUTER: IBM PC compatible
/ OPERATING SYSTEM: PC-DOS/MS-DOS
/ SOFTWARE: Patentn Release #1.0, Version #1.25
/ CURRENT APPLICATION DATA:
/ APPLICATION NUMBER: US/09/430,854
/ FILING DATE:
/ CLASSIFICATION:
/ PRIOR APPLICATION DATA:
/ APPLICATION NUMBER: 08/845,998
/ FILING DATE:
/ ATTORNEY/AGENT INFORMATION:
/ NAME: Van Amsterdam, John R.
/ REGISTRATION NUMBER: 40,212
/ REFERENCE/DOCKET NUMBER: L0461/7008
/ TELECOMMUNICATION INFORMATION:
/ TELEPHONE: (617)720-3500
/ TELEFAX: (617)720-2441
/ INFORMATION FOR SEQ ID NO: 4:
/ SEQUENCE CHARACTERISTICS:
/ LENGTH: 190 amino acids
/ TYPE: amino acid
/ TOPOLOGY: linear
/ MOLECULE TYPE: protein
/
/ US-09-430-854-4
/
/ Query Match 18.4%; Score 101.5; DB 4; Length 190;
/ Best Local Similarity 39.7%; Pred. No. 0.00012;
/ Matches 25; Conservative 10; Mismatches 19; Indels 9; Gaps 2;
/
/ QY 40 IAGTSGFDIAVLAQWVNLQSGIRVLSLPQPSDIPPGTGVFIYGYGRDNDRDPSSRK 99
/ DB 115 IIQT--GADIALLELEPVPNISHTVTLPPASETFPPGMPGVCWTGMGDVDN----- 165
/
/ QY 100 NGG 102
/ DB 166 NGG 168
/
/ RESULT 8
/ US-09-430-854-6
/ Sequence 6, Application US/09430854
/ Patent No. 6271019
/ GENERAL INFORMATION:
/ APPLICANT: Van Baren, Nicolas
/ APPLICANT: Coulie, Pierre G.
/ APPLICANT: De Smet, Charles
/ APPLICANT: Lucas, Sophie
/ APPLICANT: Boon, Thierry
/ TITLE OF INVENTION: LEUKEMIA ASSOCIATED GENES
/ NUMBER OF SEQUENCES: 16
/ CORRESPONDENCE ADDRESS:
/ ADDRESSER: Wolf, Greenfield & Sacks, P.C.
/ STREET: 600 Atlantic Avenue
/ CITY: Boston
/ STATE: MA
```

```
/ COUNTRY: US
/ ZIP: 02210
/
/ COMPUTER READABLE FORM:
/ MEDIUM TYPE: Floppy disk
/ COMPUTER: IBM PC compatible
/ OPERATING SYSTEM: PC-DOS/MS-DOS
/ SOFTWARE: Patentn Release #1.0, Version #1.25
/ CURRENT APPLICATION DATA:
/ APPLICATION NUMBER: US/09/430,854
/ FILING DATE:
/ CLASSIFICATION:
/ PRIOR APPLICATION DATA:
/ APPLICATION NUMBER: 08/845,998
/ FILING DATE:
/ ATTORNEY/AGENT INFORMATION:
/ NAME: Van Amsterdam, John R.
/ REGISTRATION NUMBER: 40,212
/ REFERENCE/DOCKET NUMBER: L0461/7008
/ TELECOMMUNICATION INFORMATION:
/ TELEPHONE: (617)720-3500
/ TELEFAX: (617)720-2441
/ INFORMATION FOR SEQ ID NO: 6:
/ SEQUENCE CHARACTERISTICS:
/ LENGTH: 190 amino acids
/ TYPE: amino acid
/ TOPOLOGY: linear
/ MOLECULE TYPE: protein
/
/ US-09-430-854-6
/
/ Query Match 18.4%; Score 101.5; DB 4; Length 190;
/ Best Local Similarity 39.7%; Pred. No. 0.00012;
/ Matches 25; Conservative 10; Mismatches 19; Indels 9; Gaps 2;
/
/ QY 40 IAGTSGFDIAVLAQWVNLQSGIRVLSLPQPSDIPPGTGVFIYGYGRDNDRDPSSRK 99
/ DB 115 IIQT--GADIALLELEPVPNISHTVTLPPASETFPPGMPGVCWTGMGDVDN----- 165
/
/ QY 100 NGG 102
/ DB 166 NGG 168
/
/ RESULT 9
/ US-08-978-404B-8
/ Sequence 8, Application US/08978404B
/ Patent No. 5968782
/ GENERAL INFORMATION:
/ APPLICANT: Stevens, Richard L.
/ TITLE OF INVENTION: MAST CELL PROTEASE THAT CLEAVES
/ NUMBER OF SEQUENCES: 74
/ CORRESPONDENCE ADDRESS:
/ ADDRESSER: Wolf, Greenfield & Sacks, P.C.
/ STREET: 600 Atlantic Avenue
/ CITY: Boston
/ STATE: MA
/ COUNTRY: U.S.A.
/ ZIP: 02210-2211
/
/ COMPUTER READABLE FORM:
/ MEDIUM TYPE: Diskette
/ COMPUTER: IBM Compatible
/ OPERATING SYSTEM: DOS
/ SOFTWARE: FastSeq for Windows Version 2.0
/ CURRENT APPLICATION DATA:
/ APPLICATION NUMBER: US/08/978,404B
/ FILING DATE: 25-NOV-97
/ CLASSIFICATION: 435
/ PRIOR APPLICATION DATA:
/ APPLICATION NUMBER: 60/032,354
/ FILING DATE: 04-DEC-1996
/ ATTORNEY/AGENT INFORMATION:
/ NAME: Plumer, Elizabeth R.
/ REGISTRATION NUMBER: 36,637
```

```

/ REFERENCE/DOCKET NUMBER: B0801/7090
/ TELECOMMUNICATION INFORMATION:
/ TELEPHONE: 617-720-3500
/ TELEFAX: 617-720-2441
/ TELEX:
/ INFORMATION FOR SEQ ID NO: 8:
/ SEQUENCE CHARACTERISTICS:
/ LENGTH: 270 amino acids
/ TYPE: amino acid
/ STRANDEDNESS: single
/ TOPOLOGY: linear
/ MOLECULE TYPE: No. 5968782e
US-08-978-404B-8

Query Match 18.3%; Score 101; DB 2; Length 270;
Best Local Similarity 41.2%; Pred. No. 0.00022;
Matches 21; Conservative 10; Mismatches 20; Indels 0; Gaps 0;

Cy 43 TLSSGDIAIWLACQVNLQSGIRVLSLPQSPDIPPGGVRIVYGGDDND 93
Db 111 TONGADIALLEKNPKNISVHVPSLPPASSTPPSGTLCTWGTGMGIND 161

RESULT 10
US-08-944-483-69
/ Sequence 69, Application US/08944483
/ Patent No. 6232456
/ GENERAL INFORMATION:
/ APPLICANT: COHEN, MAURICE
/ APPLICANT: COLPITS, TRACEY L.
/ APPLICANT: FRIEDMAN, PAULA N.
/ APPLICANT: GRANADOS, EDWARD N.
/ APPLICANT: KLASS, MICHAEL R.
/ APPLICANT: RUSSELL, JOHN C.
/ APPLICANT: STEWART, KENT D.
/ APPLICANT: STROUPE, STEVEN D.
/ TITLE OF INVENTION: NOVEL SERINE PROTEASE REAGENTS
/ TITLE OF INVENTION: AND METHODS USEFUL FOR DETECTING AND TREATING DISEASES
/ NUMBER OF INVENTION: OF THE PROSTATE
/ NUMBER OF SEQUENCES: 76
/ CORRESPONDENCE ADDRESS:
/ ADDRESSEE: Abbott Laboratories
/ STREET: 100 Abbott Park Road
/ CITY: Abbott Park
/ STATE: IL
/ COUNTRY: USA
/ ZIP: 60064-3500
/ COMPUTER READABLE FORM:
/ MEDIUM TYPE: Diskette
/ COMPUTER: IBM Compatible
/ OPERATING SYSTEM: DOS
/ SOFTWARE: FastSeq for Windows Version 2.0
/ CURRENT APPLICATION DATA:
/ APPLICATION NUMBER: US/08/944,483
/ FILING DATE:
/ CLASSIFICATION: 424
/ PRIOR APPLICATION DATA:
/ APPLICATION NUMBER:
/ FILING DATE:
/ ATTORNEY/AGENT INFORMATION:
/ NAME: Becker, Cheryl L.
/ REGISTRATION NUMBER: 35,441
/ REFERENCE/DOCKET NUMBER: 6183.US.01
/ TELECOMMUNICATION INFORMATION:
/ TELEPHONE: 847/935-1729
/ TELEFAX: 847/938-2623
/ TELEX:
/ INFORMATION FOR SEQ ID NO: 69:
/ SEQUENCE CHARACTERISTICS:
/ LENGTH: 245 amino acids
/ TYPE: amino acid
/ STRANDEDNESS: single
/ TOPOLOGY: linear

```

```

; MOLECULE TYPE: No. 6232456e
US-08-944-483-69

Query Match 18.1%; Score 100; DB 4; Length 245;
Best Local Similarity 40.7%; Pred. No. 0.00025;
Matches 22; Conservative 10; Mismatches 20; Indels 2; Gaps 1

Cy 40 IAGTISGDIATVMAQWVNLOSIGRIVSLPQSPDIPPGTGVFIVGGRDND 93
| | | | | : | : | : | : | : | : | : | : | : | : | : |
Db 85 IIQT--GADIALLELEFPVNISRVHTVMLPASETFPPGMCWVGMDVND 136

RESULT 11
US-09-016-366A-17
; Sequence 17, Application US/09016366A
; Patent No. 5955431
; GENERAL INFORMATION:
; APPLICANT: Stevens, Richard L.
; APPLICANT: Huang, Chifu
; TITLE OF INVENTION: MAST CELL PROTEASE PEPTIDE
; TITLE OF INVENTION: INHIBITORS
; NUMBER OF SEQUENCES: 65
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Wolf, Greenfield & Sacks, P.C.
; STREET: 600 Atlantic Avenue
; CITY: Boston
; STATE: MA
; COUNTRY: U.S.A.
; ZIP: 02210-2211
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/016.366A
; FILING DATE: January 30, 1998
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 60/037,090
; FILING DATE: 05-FEB-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Plumer, Elizabeth R.
; REGISTRATION NUMBER: 36,637
; REFERENCE/DOCKET NUMBER: B0801/7093
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617-720-3500
; TELEFAX: 617-720-2441
; TEXT:
; INFORMATION FOR SEQ ID NO: 17:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 275 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-09-016-366A-17

Query Match 18.1%; Score 100; DB 2; Length 275;
Best Local Similarity 40.7%; Pred. No. 0.00029;
Matches 22; Conservative 10; Mismatches 20; Indels 2; Gaps 1

Cy 40 IAGTISGDIATVMAQWVNLOSIGRIVSLPQSPDIPPGTGVFIVGGRDND 93
| | | | | : | : | : | : | : | : | : | : | : | : | : |
Db 115 IIQT--GADIALLELEFPVNISRVHTVMLPASETFPPGMCWVGMDVND 166

RESULT 12
US-08-978-404B-12
; Sequence 12, Application US/08978404B
; Patent No. 5968782
; GENERAL INFORMATION:
; APPLICANT: Stevens, Richard L.

```

```

; TITLE OF INVENTION: MAST CELL PROTEASE THAT CLEAVES
; TITLE OF INVENTION: FIBRINOGEN
; NUMBER OF SEQUENCES: 74
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Wolf, Greenfield & Sacks, P.C.
; STREET: 600 Atlantic Avenue
; CITY: Boston
; STATE: MA
; COUNTRY: U.S.A.
; ZIP: 02210-2211
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/978,404B
; FILING DATE: 25-NOV-97
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 60/032,354
; FILING DATE: 04-DEC-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Plumer, Elizabeth R.
; REGISTRATION NUMBER: 36,637
; REFERENCE/DOCKET NUMBER: B0801/7090
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617-720-3500
; TELEFAX: 617-720-2441
; TELEX:
; INFORMATION FOR SEQ ID NO: 12:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 275 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: No. 5968782e
; US-08-978-404B-12

Query Match 18.1%; Score 100; DB 2; Length 275;
Best Local Similarity 40.7%; Pred. No. 0.00029;
Matches 22; Conservative 10; Mismatches 20; Indels 2; Gaps 1;

Oy 40 IAGTSGFDIAVLAQMNLQSGIRVLSLPQSDIPPPGTGVFIYXGRDND 93
    |||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
Db 115 IIGT--GADIALLELEPYNVSHVHTVTLPPASETFPPGMPCWVTGMDVDND 166

RESULT 13
US-09-016-366A-19
; Sequence 19, Application US/09016366A
; Patent No. 5955431
; GENERAL INFORMATION:
; APPLICANT: Stevens, Richard L.
; APPLICANT: Huang, Chifu
; TITLE OF INVENTION: MAST CELL PROTEASE PEPTIDE
; TITLE OF INVENTION: INHIBITORS
; NUMBER OF SEQUENCES: 65
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Wolf, Greenfield & Sacks, P.C.
; STREET: 600 Atlantic Avenue
; CITY: Boston
; STATE: MA
; COUNTRY: U.S.A.
; ZIP: 02210-2211
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/016,366A
; FILING DATE: January 30, 1998
```

```

; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 60/037,090
; FILING DATE: 05-FEB-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Plumer, Elizabeth R.
; REGISTRATION NUMBER: 36,637
; REFERENCE/DOCKET NUMBER: B0801/7093
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617-720-3500
; TELEFAX: 617-720-2441
; TELEX:
; INFORMATION FOR SEQ ID NO: 19:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 273 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; US-09-016-366A-19

Query Match 17.9%; Score 99; DB 2; Length 273;
Best Local Similarity 39.6%; Pred. No. 0.00038;
Matches 19; Conservative 11; Mismatches 18; Indels 0; Gaps 0;

Oy 46 GFDIAVLAQMNLQSGIRVLSLPQSDIPPPGTGVFIYXGRDND 93
    |||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
Db 117 GADIALLELEPYNVSHVHTVTLPPASETFPPGMPCWVTGMDVDND 164

RESULT 14
US-08-978-404B-6
; Sequence 6, Application US/08978404B
; Patent No. 5968782
; GENERAL INFORMATION:
; APPLICANT: Stevens, Richard L.
; TITLE OF INVENTION: MAST CELL PROTEASE THAT CLEAVES
; TITLE OF INVENTION: FIBRINOGEN
; NUMBER OF SEQUENCES: 74
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Wolf, Greenfield & Sacks, P.C.
; STREET: 600 Atlantic Avenue
; CITY: Boston
; STATE: MA
; COUNTRY: U.S.A.
; ZIP: 02210-2211
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/978,404B
; FILING DATE: 25-NOV-97
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 60/032,354
; FILING DATE: 04-DEC-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Plumer, Elizabeth R.
; REGISTRATION NUMBER: 36,637
; REFERENCE/DOCKET NUMBER: B0801/7090
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617-720-3500
; TELEFAX: 617-720-2441
; TELEX:
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 273 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: No. 5968782e
; US-08-978-404B-6
```

Query Match	17.9%	Score 99	DB 2	Length 273
Best Local Similarity	40.7%	Pred. No. 0.00038		
Matches 22	Conservative 10	Mismatches 20	Indels 2	Gaps 1

QY 40 IAQTLSGFDAIVMLAQWNLQSGIRVISLPQPSDIPPPGTGVFVIGYEADDND 93  
||| ||| : | : : ||| : ||| : ||| : |||  
Db 113 IAQ--DGADIALKLTPNVNITSNVHTSLPASEETPPSGTLCWVTGWGNIND 16

RESULT 15  
US-08-978-404B-14

Query Match	17.9%	Score	99	DB	2	Length	273
Best Local Similarity	39.6%	Pred.	No.	0.00038			
Matches	19	Conservative	18	Indels	0	Gaps	0

QY 46 GPEDIAVLQAQWNLQSGIRVLSLPQPSDIPPGTGVFIIVGYGRDDND 93  
| ||| : | : | : : ||| : ||| : ||| : ||| : ||| : ||| : ||| :  
Db 117 GADIALLELEEPVNSSHVHTVTLPASSETFPPGMPCWTVGMGDVDND 164

RESULT 16  
 US-08-944-483-58  
 : Sequence 58, Application US/08944483  
 : Patent No. 6232456  
 :  
 : GENERAL INFORMATION:  
 :  
 : APPLICANT: COHEN, MAURICE  
 :  
 : APPLICANT: COLPITTS, TRACEY L.  
 :  
 : APPLICANT: FRIEDMAN, PAULA N.  
 :  
 : APPLICANT: GRANAPOS, EDWARD N.

1 APPLICANT: KLASS, MICHAEL R.  
2 APPLICANT: RUSSELL, JOHN C.  
3 APPLICANT: STEWART, KENT D.  
4 APPLICANT: STROUPE, STEVEN D.  
5 TITLE OF INVENTION: NOVEL SERINE PROTEASE REAGENTS  
6 TITLE OF INVENTION: AND METHODS USEFUL FOR DETECTING AND TREATING DISEASES  
7 OF THE PROSTATE  
8  
9

Query March	17.3%	Score 95.5;	DB 4;	Length 242;
Best Local Similarity	29.1%	Pred. NO. 0.0087;		
Matches 25; Conservative	16;	Mismatches 34;	Indels 11;	Gaps 14;

```

QY      4 L T L N G D O G I H H D P S G V K V A P G Y M S C M S A Q R R P I A Q T L S G D I A I W L A Q M V L Q S G   63
      |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db     62 L A V E G P E G V I P I N S E L F V H P I L M R N S C V A C ----- G N D I A L I K T S R S A Q L G D A   110

```

```

QY      64  IRVLSLPQPSDIPPPGTGVFIVGGR      89
          ::: ||| ||| ::|||
Db      111 VQLASLPAGDILPNKTPCYITGWGR      136

```

RESULT 17  
 US-09-079-970A-6  
 ? Sequence 6, Application US/09079970A  
 ? Patent No. 6274366  
 ? GENERAL INFORMATION:  
 ? APPLICANT: Maffitt, Mark A.  
 ? APPLICANT: Niles, Andrew L.  
 ? APPLICANT: Haak-Fendtscho, Mary  
 ? TITLE OF INVENTION: Enzymatically-Active Recombinant Human  
 ? TITLE OF INVENTION: Beta-Trypsin and Method of Making Same  
 ? NUMBER OF SEQUENCES: 6  
 ? CORRESPONDENCE ADDRESS:  
 ? ADDRESSEE: Intellectual Property Department  
 ? STREET: 8000 Excelsior Drive, Suite 401  
 ? CITY: Madison  
 ? STATE: WISCONSIN

```

; COUNTRY: U.S.A.
; ZIP: 53717-1914
; COMPUTER READABLE FORM:
; MEDIUM TYPE: floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/079,970A
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Leone, Joseph T.
; REGISTRATION NUMBER: 37,170
; REFERENCE/DOCKET NUMBER: 34506.073
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (608) 831-2100
; TELEFAX: (608) 831-2106
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 245 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-09-079-970A-6

Query Match      16.8%; Score 93; DB 4; Length 245;
Best Local Similarity 37.5%; Pred. No. 0.0018;
Matches 18; Conservative 11; Mismatches 19; Indels 0; Gaps 0;

Qy 46 GFDIAVMAQWNLQSGIRVLSLPQSDIPPGTGVFVGYGRDND 93
Db 89 GADIALLEBEPVKVSSHVHTVTLPPASETFPPGMPCMVWTGMDVDND 136

RESULT 18
US-09-079-970A-5
; Sequence 5, Application US/09079970A
; Patent No. 6274366
; GENERAL INFORMATION:
; APPLICANT: Maffitt, Mark A.
; APPLICANT: Niles, Andrew L.
; TITLE OF INVENTION: Enzymatically-Active Recombinant Human
; NUMBER OF INVENTION: Beta-Trypsin and Method of Making Same
; CORRESPONDENCE ADDRESS:
; ADDRESSER: Intellectual Property Department
; STREET: 8000 Excelsior Drive, Suite 401
; CITY: Madison
; STATE: WISCONSIN
; COUNTRY: U.S.A.
; ZIP: 53717-1914
; COMPUTER READABLE FORM:
; MEDIUM TYPE: floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/079,970A
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Leone, Joseph T.
; REGISTRATION NUMBER: 37,170
; REFERENCE/DOCKET NUMBER: 34506.073
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (608) 831-2100
; TELEFAX: (608) 831-2106
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 249 amino acids
; TYPE: amino acid
```

```

; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-09-079-970A-5

Query Match      16.8%; Score 93; DB 4; Length 249;
Best Local Similarity 37.5%; Pred. No. 0.0018;
Matches 18; Conservative 11; Mismatches 19; Indels 0; Gaps 0;

Qy 46 GFDIAVMAQWNLQSGIRVLSLPQSDIPPGTGVFVGYGRDND 93
Db 93 GADIALLEBEPVKVSSHVHTVTLPPASETFPPGMPCMVWTGMDVDND 140

RESULT 19
US-09-016-366A-23
; Sequence 23, Application US/09016366A
; Patent No. 5955431
; GENERAL INFORMATION:
; APPLICANT: Stevens, Richard L.
; APPLICANT: Huang, Chifu
; TITLE OF INVENTION: MAST CELL PROTEASE PEPTIDE
; NUMBER OF INVENTION: INHIBITORS
; CORRESPONDENCE ADDRESS:
; ADDRESSER: Wolf, Greenfield & Sacks, P.C.
; STREET: 600 Atlantic Avenue
; CITY: Boston
; STATE: MA
; COUNTRY: U.S.A.
; ZIP: 02210-2211
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/016,366A
; FILING DATE: January 30, 1998
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 60/037,090
; FILING DATE: 05-FEB-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Plummer, Elizabeth R.
; REGISTRATION NUMBER: 36,637
; REFERENCE/DOCKET NUMBER: B0801/7093
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617-720-3500
; TELEFAX: 617-720-2441
; TEXAS:
; INFORMATION FOR SEQ ID NO: 23:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 267 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; US-09-016-366A-23

Query Match      16.8%; Score 93; DB 2; Length 267;
Best Local Similarity 37.5%; Pred. No. 0.002;
Matches 18; Conservative 11; Mismatches 19; Indels 0; Gaps 0;

Qy 46 GFDIAVMAQWNLQSGIRVLSLPQSDIPPGTGVFVGYGRDND 93
Db 111 GADIALLEBEPVKVSSHVHTVTLPPASETFPPGMPCMVWTGMDVDND 158

RESULT 20
US-08-978-404B-18
; Sequence 18, Application US/08978404B
; Patent No. 5968782
; GENERAL INFORMATION:
; APPLICANT: Stevens, Richard L.
```

```

; TITLE OF INVENTION: MAST CELL PROTEASE THAT CLEAVES
; TITLE OF INVENTION: FIBRINOGEN
; NUMBER OF SEQUENCES: 74
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Wolf, Greenfield & Sacks, P.C.
; STREET: 600 Atlantic Avenue
; CITY: Boston
; STATE: MA
; COUNTRY: U.S.A.
; ZIP: 02210-2211
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/978,404B
; FILING DATE: 25-NOV-97
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 60/032,354
; FILING DATE: 04-DEC-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Plumer, Elizabeth R.
; REGISTRATION NUMBER: 36,637
; REFERENCE/DOCKET NUMBER: B0801/7090
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617-720-3500
; TELEFAX: 617-720-2441
; TELEX:
; INFORMATION FOR SEQ ID NO: 18:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 267 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: No. 5968782e
; US-08-978-404B-18

Query Match 16.8%; Score 93; DB 2; Length 267;
Best Local Similarity 37.5%; Pred. No. 0.002;
Matches 18; Conservative 11; Mismatches 19; Indels 0; Gaps 0;

Qy 46 GFDIAVMAQWVNLQSGIRVLSLPQSDIPPPGTGTFVIGYGRDND 93
Db 111 GADIALLEBEPVKVSHVHTVTLPPASETFPPGKPCWVTGMDVDND 158

RESULT 21
US-09-016-366A-21
; Sequence 21, Application US/09016366A
; Patent No. 5955431
; GENERAL INFORMATION:
; APPLICANT: Stevens, Richard L.
; APPLICANT: Huang, Chifu
; TITLE OF INVENTION: MAST CELL PROTEASE PEPTIDE
; TITLE OF INVENTION: INHIBITORS
; NUMBER OF SEQUENCES: 65
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Wolf, Greenfield & Sacks, P.C.
; STREET: 600 Atlantic Avenue
; CITY: Boston
; STATE: MA
; COUNTRY: U.S.A.
; ZIP: 02210-2211
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/016,366A
; FILING DATE: January 30, 1998
```

```

; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 60/037,090
; FILING DATE: 05-FEB-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Plumer, Elizabeth R.
; REGISTRATION NUMBER: 36,637
; REFERENCE/DOCKET NUMBER: B0801/7093
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617-720-3500
; TELEFAX: 617-720-2441
; TELEX:
; INFORMATION FOR SEQ ID NO: 21:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 274 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; US-09-016-366A-21

Query Match 16.8%; Score 93; DB 2; Length 274;
Best Local Similarity 37.5%; Pred. No. 0.002;
Matches 18; Conservative 11; Mismatches 19; Indels 0; Gaps 0;

Qy 46 GFDIAVMAQWVNLQSGIRVLSLPQSDIPPPGTGTFVIGYGRDND 93
Db 118 GADIALLEBEPVKVSHVHTVTLPPASETFPPGKPCWVTGMDVDND 165

RESULT 22
US-08-978-404B-16
; Sequence 16, Application US/08978404B
; Patent No. 5968782
; GENERAL INFORMATION:
; APPLICANT: Stevens, Richard L.
; TITLE OF INVENTION: MAST CELL PROTEASE THAT CLEAVES
; TITLE OF INVENTION: FIBRINOGEN
; NUMBER OF SEQUENCES: 74
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Wolf, Greenfield & Sacks, P.C.
; STREET: 600 Atlantic Avenue
; CITY: Boston
; STATE: MA
; COUNTRY: U.S.A.
; ZIP: 02210-2211
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/978,404B
; FILING DATE: 25-NOV-97
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 60/032,354
; FILING DATE: 04-DEC-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Plumer, Elizabeth R.
; REGISTRATION NUMBER: 36,637
; REFERENCE/DOCKET NUMBER: B0801/7090
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617-720-3500
; TELEFAX: 617-720-2441
; TELEX:
; INFORMATION FOR SEQ ID NO: 16:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 274 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: No. 5968782e
; US-08-978-404B-16
```





```

; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/978.404B
; FILING DATE: 25-NOV-97
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 60/032,354
; FILING DATE: 04-DEC-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Plumer, Elizabeth R.
; REGISTRATION NUMBER: 36,637
; REFERENCE/DOCKET NUMBER: B0801/7090
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617-720-3500
; TELEFAX: 617-720-2441
;
; TELEX:
;
; INFORMATION FOR SEQ ID NO: 10:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 269 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: No. 5968782e
;
; US-08-978-404B-10

```

```

Query Match 15.1%; Score 85.5; DB 2; Length 269;
Best Local Similarity 38.8%; Pred. No. 0.016;
Matches 19; Conservative 9; Mismatches 20; Indels 1; Gaps 1;

```

```

OY 48 DIAIYMLQWNLQSGIRVISLPPSPDIPPGTGVPIVYGGDDMDRDP 96
DB 116 DIAIYMLQWNLQSGIRVISLPPSPDIPPGTGVPIVYGGDDMDRDP 163

```

```

RESULT 26
US-09-386-653A-7
; Sequence 7, Application US/09386653A
; Patent No. 6458564
; GENERAL INFORMATION:
; APPLICANT: Andrade-Gordon, Patricia
; APPLICANT: Darrow, Andrew
; APPLICANT: Qi, Jian-shen
; TITLE OF INVENTION: DNA encoding the novel human serine
; FILE REFERENCE: ORT-1032
; CURRENT APPLICATION NUMBER: US/09/386,653A
; CURRENT FILING DATE: 1999-08-31
; NUMBER OF SEQ ID NOS: 11
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 7
; LENGTH: 290
; TYPE: PRT
; ORGANISM: Homo sapiens
; US-09-386-653A-7

```

```

Query Match 15.1%; Score 83.5; DB 4; Length 290;
Best Local Similarity 32.6%; Pred. No. 0.03;
Matches 29; Conservative 13; Mismatches 38; Indels 9; Gaps 4;

```

```

OY 7 RNDGQGGIHHQPSGVK--VAPGYMSCSARQR-----PIAQ-TISGPIAIVMLQWYN 59
DB 78 RNTSETSLYQVLLGARQLVQPG--PHANYARVRQVESNPLYQGTASADVALVELLEAPVP 135
OY 60 LQSGIRVISLPPSPDIPPGTGVPIVYGG 88
DB 136 FTNYILPVCLDPSPVIFETGMNCWVTGMG 164

```

RESULT 27

```

US-09-386-653A-9
; Sequence 9, Application US/09386653A
; Patent No. 6458564
; GENERAL INFORMATION:
; APPLICANT: Andrade-Gordon, Patricia
; APPLICANT: Darrow, Andrew
; APPLICANT: Qi, Jian-shen
; TITLE OF INVENTION: DNA encoding the novel human serine
; FILE REFERENCE: ORT-1032
; CURRENT APPLICATION NUMBER: US/09/386,653A
; CURRENT FILING DATE: 1999-08-31
; NUMBER OF SEQ ID NOS: 11
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 9
; LENGTH: 315
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Fusion Protein
; US-09-386-653A-9

```

```

Query Match 15.1%; Score 83.5; DB 4; Length 315;
Best Local Similarity 32.6%; Pred. No. 0.034;
Matches 29; Conservative 13; Mismatches 38; Indels 9; Gaps 4;

```

```

OY 7 RNDGQGGIHHQPSGVK--VAPGYMSCSARQR-----PIAQ-TISGPIAIVMLQWYN 59
DB 95 RNTSETSLYQVLLGARQLVQPG--PHANYARVRQVESNPLYQGTASADVALVELLEAPVP 152
OY 60 LQSGIRVISLPPSPDIPPGTGVPIVYGG 88
DB 153 FTNYILPVCLDPSPVIFETGMNCWVTGMG 181

```

```

RESULT 28
US-08-978-404B-3
; Sequence 3, Application US/08978404B
; Patent No. 5968782
; GENERAL INFORMATION:
; APPLICANT: Stevens, Richard L.
; TITLE OF INVENTION: MAST CELL PROTEASE THAT CLEAVES
; NUMBER OF SEQUENCES: 74
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Wolf, Greenfield & Sacks, P.C.
; STREET: 600 Atlantic Avenue
; CITY: Boston
; STATE: MA
; COUNTRY: U.S.A.
; ZIP: 02210-2211
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/978.404B
; FILING DATE: 25-NOV-97
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 60/032,354
; FILING DATE: 04-DEC-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Plumer, Elizabeth R.
; REGISTRATION NUMBER: 36,637
; REFERENCE/DOCKET NUMBER: B0801/7090
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617-720-3500
; TELEFAX: 617-720-2441
; TELEX:
; INFORMATION FOR SEQ ID NO: 3:

```

SEQUENCE CHARACTERISTICS:  
LENGTH: 273 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: No. 5968782e  
US-08-978-404B-3

Query Match 15.0%; Score 83; DB 2; Length 273;  
Best Local Similarity 36.2%; Pred. No. 0.03;  
Matches 17; Conservative 9; Mismatches 21; Indels 0; Gaps 0;

OY 46 GFDIAIWLQWVNLQSGIRVLSLPQPSDIPPGTGVFIYGYGRDN 92  
DB 117 GADIALKLKLTNPVNTSDYVHPVPLPASPSTFPGTLCWVTGNGNIDN 163

## RESULT 29

US-09-027-337-2  
Sequence 2, Application US/09027337B  
Patent No. 5972616  
GENERAL INFORMATION:  
APPLICANT: O'Brien, Timothy J.  
APPLICANT: Tanimoto, Hirotsoshi  
TITLE OF INVENTION: TADG-15: An Extracellular Serine Protease Overexpressed in  
FILE REFERENCE: D6064  
CURRENT APPLICATION NUMBER: US/09/027,337B  
NUMBER OF SEQ ID NOS: 13  
SEQ ID NO 2  
LENGTH: 855  
TYPE: PRT  
ORGANISM: Homo sapiens  
FEATURE:  
OTHER INFORMATION: Amino acid sequence of TADG-15 encoded by nucleotides  
Patent No. 5972616  
US-09-027-337-2

Query Match 14.9%; Score 82.5; DB 2; Length 855;  
Best Local Similarity 27.1%; Pred. No. 0.16;  
Matches 23; Conservative 15; Mismatches 28; Indels 19; Gaps 3;

OY 13 GIHHQPSGVKVA PGWPSGMSARORRPIAQTLS-----GFDIAIWLQWVNLQSG 63  
DB 677 GIHHQSQ--RSAPGV-----QERRLKRIISHPFNDFTPDYDIALLEKPAEYSSM 726

OY 64 IRVLSLPQPSDIPPGTGVFIYGYG 88  
DB 727 VRPICLPDASHVFPAGKAIWVGWG 751

## RESULT 30

US-09-644-600-2  
Sequence 2, Application US/09644600  
Patent No. 6451500  
GENERAL INFORMATION:  
APPLICANT: O'Brien, Timothy J.  
APPLICANT: Tanimoto, Hirotsoshi  
TITLE OF INVENTION: TADG-15: An Extracellular Serine Protease  
FILE REFERENCE: D6064CIP/D  
CURRENT APPLICATION NUMBER: US/09/644,600  
PRIOR FILING DATE: 2000-08-23  
PRIOR APPLICATION NUMBER: 09/421,213  
PRIOR FILING DATE: 1999-10-20  
PRIOR APPLICATION NUMBER: 09/027,337  
NUMBER OF SEQ ID NOS: 98  
SEQ ID NO 2  
LENGTH: 855  
TYPE: PRT

ORGANISM: Homo sapiens  
FEATURE:  
OTHER INFORMATION: TADG-15  
US-09-644-600-2

Query Match 14.9%; Score 82.5; DB 4; Length 855;  
Best Local Similarity 27.1%; Pred. No. 0.16;  
Matches 23; Conservative 15; Mismatches 28; Indels 19; Gaps 3;

OY 13 GIHHQPSGVKVA PGWPSGMSARORRPIAQTLS-----GFDIAIWLQWVNLQSG 63  
DB 677 GIHHQSQ--RSAPGV-----QERRLKRIISHPFNDFTPDYDIALLEKPAEYSSM 726

OY 64 IRVLSLPQPSDIPPGTGVFIYGYG 88  
DB 727 VRPICLPDASHVFPAGKAIWVGWG 751

## RESULT 31

US-08-278-091-12  
Sequence 12, Application US/08278091  
Patent No. 5506139  
GENERAL INFORMATION:  
APPLICANT: LOOSMORE, Sheena M  
APPLICANT: YANG, Yan-Ping  
APPLICANT: CHONG, Pele  
APPLICANT: COHEN, Raymond P.  
APPLICANT: KLEIN, Michel H.  
TITLE OF INVENTION: Analog of Haemophilus Hinf7 Protein with  
NUMBER OF SEQUENCES: 23  
CORRESPONDENCE ADDRESS:  
ADDRESS: Slim & McBurney  
STREET: Suite 701, 330 University Avenue  
CITY: Toronto  
STATE: Ontario  
COUNTRY: Canada  
ZIP: M5G 1R7  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/278,091  
FILING DATE: 21-JUL-1994  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Stewart, Michael I  
REGISTRATION NUMBER: 24,973  
REFERENCE/DOCKET NUMBER: 1038-371  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (416) 595-1155  
TELEFAX: (416) 595-1163  
INFORMATION FOR SEQ ID NO: 12:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 224 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
US-08-278-091-12

Query Match 14.8%; Score 81.5; DB 1; Length 224;  
Best Local Similarity 36.0%; Pred. No. 0.038;  
Matches 18; Conservative 9; Mismatches 22; Indels 1; Gaps 1;

OY 48 DIAIWLQWVNLQSGIRVLSLPQPSDIPPGTGVFIYGYGRDNDRDP 97  
DB 89 DIMLKLEKVELTPAVNVVPLPSPSDFIHGAMCMAAGWGR-TGVDRPT 137

## RESULT 32

US-08-483-859-12

```

; Sequence 12, Application US/08483859
; Patent No. 5656436
; GENERAL INFORMATION:
; APPLICANT: LOOSMORE, Sheena M.
; APPLICANT: YANG, Yan-Ping
; APPLICANT: CHONG, Pele
; APPLICANT: COHEN, Raymond P.
; APPLICANT: KLEIN, Michael H.
; TITLE OF INVENTION: Analog of Haemophilus Hinf7 Protein with
; TITLE OF INVENTION: Reduced Protease Activity
; NUMBER OF SEQUENCES: 23
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sim & Mcburney
; STREET: Suite 701, 330 University Avenue
; CITY: Toronto
; STATE: Ontario
; COUNTRY: Canada
; ZIP: M5G 1R7
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/483,859
; FILING DATE: 07-JUN-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/296,149
; FILING DATE: 26-AUG-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/278,091
; FILING DATE: 21-JUL-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Stewart, Michael I.
; REGISTRATION NUMBER: 24,973
; REFERENCE/DOCKET NUMBER: 1038-495 MIS:V9
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (416) 595-1155
; TELEFAX: (416) 595-1163
; INFORMATION FOR SEQ ID NO: 12:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 224 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; US-08-483-859-12

Query Match 14.8%; Score 81.5; DB 1; Length 224;
Best Local Similarity 36.0%; Pred. No. 0.038;
Matches 18; Conservative 9; Mismatches 22; Indels 1; Gaps 1;

Qy 48 DIAIYMLAQMNVLOSIGRIVISLPQSDIPPPGTGVFIYGYGDDNDRDPS 97
Db 89 DIMIKLEKKVYELTPAVNVVPLPSPSPDPIHFGAMCWAAGMK-TGVDRPT 137

RESULT 33
US-08-472-173-12
; Sequence 12, Application US/08472173
; Patent No. 5665353
; GENERAL INFORMATION:
; APPLICANT: LOOSMORE, Sheena M
; APPLICANT: YANG, Yan-Ping
; APPLICANT: CHONG, Pele
; APPLICANT: COHEN, Raymond P.
; APPLICANT: KLEIN, Michael H.
; TITLE OF INVENTION: Analog of Haemophilus Hinf7 Protein with
; TITLE OF INVENTION: Reduced Protease Activity
; NUMBER OF SEQUENCES: 23
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sim & Mcburney
; STREET: Suite 701, 330 University Avenue

```

```

; CITY: Toronto
; STATE: Ontario
; COUNTRY: Canada
; ZIP: M5G 1R7
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/472,173
; FILING DATE: 07-JUN-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/296,149
; FILING DATE: 26-AUG-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/278,091
; FILING DATE: 21-JUL-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Stewart, Michael I.
; REGISTRATION NUMBER: 24,973
; REFERENCE/DOCKET NUMBER: 1038-493 MIS:V9
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (416) 595-1155
; TELEFAX: (416) 595-1163
; INFORMATION FOR SEQ ID NO: 12:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 224 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; US-08-472-173-12

Query Match 14.8%; Score 81.5; DB 1; Length 224;
Best Local Similarity 36.0%; Pred. No. 0.038;
Matches 18; Conservative 9; Mismatches 22; Indels 1; Gaps 1;

Qy 48 DIAIYMLAQMNVLOSIGRIVISLPQSDIPPPGTGVFIYGYGDDNDRDPS 97
Db 89 DIMIKLEKKVYELTPAVNVVPLPSPSPDPIHFGAMCWAAGMK-TGVDRPT 137

RESULT 34
US-08-487-167-12
; Sequence 12, Application US/08487167
; Patent No. 5669302
; GENERAL INFORMATION:
; APPLICANT: LOOSMORE, Sheena M.
; APPLICANT: YANG, Yan-Ping
; APPLICANT: CHONG, Pele
; APPLICANT: COHEN, Raymond P.
; APPLICANT: KLEIN, Michael H.
; TITLE OF INVENTION: Analog of Haemophilus Hinf7 Protein with
; TITLE OF INVENTION: Reduced Protease Activity
; NUMBER OF SEQUENCES: 23
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sim & Mcburney
; STREET: Suite 701, 330 University Avenue
; CITY: Toronto
; STATE: Ontario
; COUNTRY: Canada
; ZIP: M5G 1R7
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/487,167
; FILING DATE: 07-JUN-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:

```



```

: APPLICANT: YANG, Yan-Ping
: APPLICANT: CHONG, Pele
: APPLICANT: OOMEN, Raymond P.
: APPLICANT: KLEIN, Michel H.
: TITLE OF INVENTION: Analog of Haemophilus Hinf7 Protein with Reduced Protease Act
: NUMBER OF SEQUENCES: 23
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Sim & McBurney
: STREET: 6th Floor, 330 University Avenue
: CITY: Toronto
: STATE: Ontario
: COUNTRY: Canada
: ZIP: M5G 1R7
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patent Release #1.0, Version #1.25
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/801,499
: FILING DATE:
: CLASSIFICATION: 514
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 08/482,816
: FILING DATE: 07-JUN-1995
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 08/296,149
: FILING DATE: 26-AUG-1994
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 08/278,091
: FILING DATE: 21-JUL-1994
: ATTORNEY/AGENT INFORMATION:
: NAME: Stewart, Michael I
: REGISTRATION NUMBER: 24,973
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (416) 595-1155
: TELEFAX: (416) 595-1163
: INFORMATION FOR SEQ ID NO: 12:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 224 amino acids
: TYPE: amino acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: US-08-801-499-12

Query Match 14.8%; Score 81.5; DB 2; Length 224;
Best Local Similarity 36.0%; Pred. No. 0.038;
Matches 18; Conservative 9; Mismatches 22; Indels 1; Gaps 1;

Oy 48 DIAIYMLAQMVNLOSIGIRVISLPPQSDIPPGTGVFIYGRDNDNDPDS 97
Db 89 DIMLKKEKVELTPAVNVVLPSPSPDIHFGAMCMAAGMK-TGVDRPT 137

RESULT 38
: US-08-615-271-12
: Sequence 12, Application US/08615271
: Patent No. 5981503
: GENERAL INFORMATION:
: APPLICANT: LOOSMORE, Sheena M.
: APPLICANT: YANG, Yan-Ping
: APPLICANT: CHONG, Pele
: APPLICANT: OOMEN, Raymond P.
: APPLICANT: KLEIN, Michel H.
: TITLE OF INVENTION: ANALOG OF HAEMOPHILUS HINF7 WITH REDUCED
: NUMBER OF SEQUENCES: 23
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Sim & McBurney
: STREET: 6th Floor, 330 University Avenue
: CITY: Toronto
: STATE: Ontario

```

```

: COUNTRY: Canada
: ZIP: M5G 1R7
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patent Release #1.0, Version #1.30
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/615,271
: FILING DATE: 20-JUN-1996
: CLASSIFICATION: 514
: ATTORNEY/AGENT INFORMATION:
: NAME: Stewart, Michael I
: REGISTRATION NUMBER: 24,973
: REFERENCE/DOCKET NUMBER: 1038-580
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (416) 595-1155
: TELEFAX: (416) 595-1163
: INFORMATION FOR SEQ ID NO: 12:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 224 amino acids
: TYPE: amino acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: US-08-615-271-12

Query Match 14.8%; Score 81.5; DB 2; Length 224;
Best Local Similarity 36.0%; Pred. No. 0.038;
Matches 18; Conservative 9; Mismatches 22; Indels 1; Gaps 1;

Oy 48 DIAIYMLAQMVNLOSIGIRVISLPPQSDIPPGTGVFIYGRDNDNDPDS 97
Db 89 DIMLKKEKVELTPAVNVVLPSPSPDIHFGAMCMAAGMK-TGVDRPT 137

RESULT 39
: US-09-074-660-12
: Sequence 12, Application US/09074660
: Patent No. 6020183
: GENERAL INFORMATION:
: APPLICANT: LOOSMORE, Sheena M.
: APPLICANT: YANG, Yan-Ping
: APPLICANT: CHONG, Pele
: APPLICANT: OOMEN, Raymond P.
: APPLICANT: KLEIN, Michel H.
: TITLE OF INVENTION: Analog of Haemophilus Hinf7 Protein with
: NUMBER OF SEQUENCES: 23
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Sim & McBurney
: STREET: 6th Floor, 330 University Avenue
: CITY: Toronto
: STATE: Ontario
: COUNTRY: Canada
: ZIP: M5G 1R7
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patent Release #1.0, Version #1.25
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/09/074,660
: FILING DATE:
: CLASSIFICATION:
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 08/487,167
: FILING DATE: 07-JUN-1995
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 08/296,149
: FILING DATE: 26-AUG-1994
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 08/278,091
: FILING DATE: 21-JUL-1994

```

```
; ATTORNEY/AGENT INFORMATION:
; NAME: Stewart, Michael I.
; REGISTRATION NUMBER: 24,973
; REFERENCE/DOCKET NUMBER: 1038-731 MIS:jb
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (416) 595-1155
; TELEFAX: (416) 595-1163
; INFORMATION FOR SEQ ID NO: 12:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 224 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
;
US-09-074-660-12

Query Match          14.8%; Score 81.5; DB 3; Length 224;
Best Local Similarity 36.0%; Pred. No. 0.038;
Matches 18; Conservative 9; Mismatches 22; Indels 1; Gaps 1;

Qy 48 DIAIVMLAQMVNLSGIRVISLPQPSDIPPGTGVFIYVYGRDNDNDPDS 97
    ||::||:||||:||||:||||:||||:||||:||||:||||:||||:
Db 89 DIMLKLEKKVELTPAVNVVPLPSPSDFIHGAMCMAGWGK-TGVRDPT 137

RESULT 40
US-09-074-659-12
; Sequence 12, Application US/09074659
; Patent No. 6025342
; GENERAL INFORMATION:
; APPLICANT: LOOSMORE, Sheena M.
; APPLICANT: YANG, Yan-Ping
; APPLICANT: CHONG, Pele
; APPLICANT: OOMEN, Raymond P.
; APPLICANT: KLEIN, Michel H.
; TITLE OF INVENTION: Analog of Haemophilus Hina7 Protein with
; NUMBER OF SEQUENCES: 23
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Slim & McBurney
; STREET: 6th Floor, 330 University Avenue
; CITY: Toronto
; STATE: Ontario
; COUNTRY: Canada
; ZIP: M5G 1R7
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/074,659
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/487,167
; FILING DATE: 07-JUN-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/296,149
; FILING DATE: 26-AUG-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/278,091
; FILING DATE: 21-JUL-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Stewart, Michael I.
; REGISTRATION NUMBER: 24,973
; REFERENCE/DOCKET NUMBER: 1038-730 MIS:jb
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (416) 595-1155
; TELEFAX: (416) 595-1163
; INFORMATION FOR SEQ ID NO: 12:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 224 amino acids
; TYPE: amino acid
```

```
; STRANDEDNESS: single
; TOPOLOGY: linear
;
US-09-074-659-12

Query Match          14.8%; Score 81.5; DB 3; Length 224;
Best Local Similarity 36.0%; Pred. No. 0.038;
Matches 18; Conservative 9; Mismatches 22; Indels 1; Gaps 1;

Qy 48 DIAIVMLAQMVNLSGIRVISLPQPSDIPPGTGVFIYVYGRDNDNDPDS 97
    ||::||:||||:||||:||||:||||:||||:||||:||||:||||:
Db 89 DIMLKLEKKVELTPAVNVVPLPSPSDFIHGAMCMAGWGK-TGVRDPT 137

Search completed: April 9, 2003, 12:59:43
Job time : 35.0164 secs
```

GenCore version 5.1.4 p5 4578  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 9, 2003, 12:58:12 ; Search time 39.0984 Seconds  
(without alignments)  
165.747 Million cell updates/sec

Title: US-10-020-441-2

Perfect score: 552  
Sequence: 1 VSFLLRLNGDQGGIHQPSG.....YGRDNDPRSKNGILKK 106

Scoring table: BIOSUM62  
Gapop 10.0, Gapext 0.5

Searched: 248812 seqs, 61136040 residues

Total number of hits satisfying chosen parameters: 248812

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database :

Published Applications\_AA:\*  
1: /cgn2\_6/ptodata/2/pubpaa/US08\_NEW\_PUB pep:\*  
2: /cgn2\_6/ptodata/2/pubpaa/PCT\_NEW\_PUB pep:\*  
3: /cgn2\_6/ptodata/2/pubpaa/US06\_NEW\_PUB pep:\*  
4: /cgn2\_6/ptodata/2/pubpaa/US07\_PUBCOMB pep:\*  
5: /cgn2\_6/ptodata/2/pubpaa/US07\_NEW\_PUB pep:\*  
6: /cgn2\_6/ptodata/2/pubpaa/PCTUS\_PUBCOMB pep:\*  
7: /cgn2\_6/ptodata/2/pubpaa/US08\_PUBCOMB pep:\*  
8: /cgn2\_6/ptodata/2/pubpaa/US09\_NEW\_PUB pep:\*  
9: /cgn2\_6/ptodata/2/pubpaa/US09\_PUBCOMB pep:\*  
10: /cgn2\_6/ptodata/2/pubpaa/US09\_PUBCOMB pep:\*  
11: /cgn2\_6/ptodata/2/pubpaa/US10\_NEW\_PUB pep:\*  
12: /cgn2\_6/ptodata/2/pubpaa/US10\_PUBCOMB pep:\*  
13: /cgn2\_6/ptodata/2/pubpaa/US60\_NEW\_PUB pep:\*  
14: /cgn2\_6/ptodata/2/pubpaa/US60\_PUBCOMB pep:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	552	100.0	106	9	US-10-020-441-2
2	102	18.5	276	9	US-10-117-323-35
3	100	18.1	275	9	US-10-117-323-36
4	99	17.9	272	9	US-10-117-323-37
5	99	17.9	272	9	US-10-117-323-40
6	95.5	17.3	270	10	US-09-923-779-152
7	95.5	17.3	273	10	US-09-925-297-695
8	90	16.3	16	9	US-10-020-441-3
9	84	15.2	1128	10	US-09-888-615-97
10	83.5	15.1	290	9	US-10-041-006A-7
11	83.5	15.1	290	9	US-10-028-072-222
12	83.5	15.1	290	9	US-10-121-049-222
13	83.5	15.1	290	9	US-10-123-904-222
14	83.5	15.1	290	9	US-10-140-470-222
15	83.5	15.1	290	9	US-10-175-746-222
16	83.5	15.1	290	9	US-10-176-918-222
17	83.5	15.1	290	9	US-10-176-921-222
18	83.5	15.1	290	9	US-10-137-865-222
19	83.5	15.1	290	9	US-10-140-474-222

20	83.5	15.1	290	9	US-10-142-431-222	Sequence 222, App
21	83.5	15.1	290	9	US-10-143-114-222	Sequence 222, App
22	83.5	15.1	290	9	US-10-140-002-222	Sequence 222, App
23	83.5	15.1	290	9	US-10-142-419-222	Sequence 222, App
24	83.5	15.1	290	9	US-10-123-262-222	Sequence 222, App
25	83.5	15.1	290	9	US-10-142-423-222	Sequence 222, App
26	83.5	15.1	290	9	US-10-121-050-222	Sequence 222, App
27	83.5	15.1	290	9	US-10-141-755-222	Sequence 222, App
28	83.5	15.1	290	9	US-10-143-032-222	Sequence 222, App
29	83.5	15.1	290	12	US-10-040-655-7	Sequence 7, Appli
30	83.5	15.1	315	9	US-10-041-006A-9	Sequence 9, Appli
31	83.5	15.1	315	12	US-10-040-655-9	Sequence 4, Appli
32	82.5	14.9	241	9	US-10-099-700A-4	Sequence 2, Appli
33	82.5	14.9	241	9	US-10-092-004A-2	Sequence 2, Appli
34	82.5	14.9	855	9	US-10-099-700A-2	Sequence 4, Appli
35	80	14.5	279	10	US-09-832-197-4	Sequence 7, Appli
36	80	14.5	284	12	US-10-041-400A-7	Sequence 7, Appli
37	80	14.5	284	12	US-10-041-264A-7	Sequence 7, Appli
38	80	14.5	284	12	US-10-042-091A-7	Sequence 9, Appli
39	80	14.5	316	12	US-10-041-400A-9	Sequence 9, Appli
40	80	14.5	316	12	US-10-041-264A-9	Sequence 9, Appli
41	80	14.5	316	12	US-10-042-091A-9	Sequence 132, App
42	79	14.3	571	9	US-09-978-295A-132	Sequence 132, App
43	79	14.3	571	9	US-09-978-697-132	Sequence 132, App
44	79	14.3	571	9	US-09-978-192A-132	Sequence 132, App
45	79	14.3	571	9	US-09-999-832A-132	Sequence 132, App

#### ALIGNMENTS

```
RESULT 1
US-10-020-441-2
; Sequence 2, Application US/10020441
; Publication No. US200201822241
; GENERAL INFORMATION:
; APPLICANT: Doenhoff, Michael
; TITLE OF INVENTION: SCHISTOSOMIASIS VACCINE
; FILE REFERENCE: Doenhoff
; CURRENT APPLICATION NUMBER: US/10/020,441
; PRIOR FILING DATE: 2001-12-18
; PRIOR APPLICATION NUMBER: US/09/413,810
; PRIOR FILING DATE: PRIOR FILING DATE: 1999-10-07
; NUMBER OF SEQ ID NOS: 3
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2
; LENGTH: 106
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: protein
US-10-020-441-2

Query Match      100.0%; Score 552; DB 9; Length 106;
Best Local Similarity 100.0%; Pred. No. 2.4e-54;
Matches 106; Conservative 0; Mismatches 0; Indels 0;

QY 1 VSFLLRLNGDQGGIHQPSGVKVAAPGYWPCMSARQRRPIAQTLSGFDIAIVMLAQMVL 60
Db 1 VSFLLRLNGDQGGIHQPSGVKVAAPGYWPCMSARQRRPIAQTLSGFDIAIVMLAQMVL 60
QY 61 QSGRIVSLPPSPDIPPGTGVFIVGYGRDNDPRSKNGILKK 106
Db 61 QSGRIVSLPPSPDIPPGTGVFIVGYGRDNDPRSKNGILKK 106

RESULT 2
US-10-117-323-35
; Sequence 35, Application US/10117323
; Publication No. US20030054993A1
; GENERAL INFORMATION:
; APPLICANT: Rancourt, Derrick B.
```

```
; APPLICANT: Rancourt, Susan L.
; APPLICANT: O'Sullivan, Colleen M.
; TITLE OF INVENTION: Implantation Serine Proteinases
; FILE REFERENCE: 03337-005
; CURRENT APPLICATION NUMBER: US/10/117,323
; CURRENT FILING DATE: 2002-04-08
; PRIOR APPLICATION NUMBER: US 60/281,724
; PRIOR FILING DATE: 2001-04-06
; PRIOR APPLICATION NUMBER: US 60/294,736
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: US 60/350,962
; PRIOR FILING DATE: 2002-01-25
; NUMBER OF SEQ ID NOS: 41
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 35
; LENGTH: 276
; TYPE: PRT
; ORGANISM: Mouse
; US-10-117-323-35
```

```
Query Match 18.5%; Score 102; DB 9; Length 276;
Best Local Similarity 43.8%; Pred. No. 0.001;
Matches 21; Conservative 9; Mismatches 18; Indels 0; Gaps 0;
```

```
Qy 46 GFDIAIWMIAQWNLQSGIRVSLPQPSDIPPGTGVFIYGRDND 93
Db 120 GADVALLLEEVVNVSTHPIPLPASETFFPGTSCWVGMDVND 167
```

```
RESULT 3
US-10-117-323-36
; Sequence 36, Application US/10117323
; Publication No. US20030054993A1
; GENERAL INFORMATION:
; APPLICANT: Rancourt, Derrick E.
; APPLICANT: Rancourt, Susan L.
; APPLICANT: O'Sullivan, Colleen M.
; TITLE OF INVENTION: Implantation Serine Proteinases
; FILE REFERENCE: 03337-005
; CURRENT APPLICATION NUMBER: US/10/117,323
; CURRENT FILING DATE: 2002-04-08
; PRIOR APPLICATION NUMBER: US 60/281,724
; PRIOR FILING DATE: 2001-04-06
; PRIOR APPLICATION NUMBER: US 60/294,736
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: US 60/350,962
; PRIOR FILING DATE: 2002-01-25
; NUMBER OF SEQ ID NOS: 41
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 36
; LENGTH: 275
; TYPE: PRT
; ORGANISM: Homo sapiens
; US-10-117-323-36
```

```
Query Match 18.1%; Score 100; DB 9; Length 275;
Best Local Similarity 40.7%; Pred. No. 0.0017;
Matches 22; Conservative 10; Mismatches 20; Indels 2; Gaps 1;
```

```
Qy 40 IAQTISGFDIAIWMIAQWNLQSGIRVSLPQPSDIPPGTGVFIYGRDND 93
Db 115 IIQT--GADIALLEEVVNVSTHPIPLPASETFFPGMPCWVGMDVND 166
```

```
RESULT 4
US-10-117-323-37
; Sequence 37, Application US/10117323
; Publication No. US20030054993A1
; GENERAL INFORMATION:
; APPLICANT: Rancourt, Derrick E.
; APPLICANT: Rancourt, Susan L.
; APPLICANT: O'Sullivan, Colleen M.
; TITLE OF INVENTION: Implantation Serine Proteinases
```

```
; FILE REFERENCE: 03337-005
; CURRENT APPLICATION NUMBER: US/10/117,323
; CURRENT FILING DATE: 2002-04-08
; PRIOR APPLICATION NUMBER: US 60/281,724
; PRIOR FILING DATE: 2001-04-06
; PRIOR APPLICATION NUMBER: US 60/294,736
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: US 60/350,962
; PRIOR FILING DATE: 2002-01-25
; NUMBER OF SEQ ID NOS: 41
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 37
; LENGTH: 272
; TYPE: PRT
; ORGANISM: Mouse
; US-10-117-323-37
```

```
Query Match 17.9%; Score 99; DB 9; Length 272;
Best Local Similarity 40.7%; Pred. No. 0.0022;
Matches 22; Conservative 10; Mismatches 20; Indels 2; Gaps 1;
```

```
Qy 40 IAQTISGFDIAIWMIAQWNLQSGIRVSLPQPSDIPPGTGVFIYGRDND 93
Db 112 IAQ--DGADIALKLKLTNPVNTSVHTVSLPASETFFGTLCTWVGMINND 163
```

```
RESULT 5
US-10-117-323-40
; Sequence 40, Application US/10117323
; Publication No. US20030054993A1
; GENERAL INFORMATION:
; APPLICANT: Rancourt, Derrick E.
; APPLICANT: Rancourt, Susan L.
; APPLICANT: O'Sullivan, Colleen M.
; TITLE OF INVENTION: Implantation Serine Proteinases
; FILE REFERENCE: 03337-005
; CURRENT APPLICATION NUMBER: US/10/117,323
; CURRENT FILING DATE: 2002-04-08
; PRIOR APPLICATION NUMBER: US 60/281,724
; PRIOR FILING DATE: 2001-04-06
; PRIOR APPLICATION NUMBER: US 60/294,736
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: US 60/350,962
; PRIOR FILING DATE: 2002-01-25
; NUMBER OF SEQ ID NOS: 41
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 40
; LENGTH: 272
; TYPE: PRT
; ORGANISM: Mouse
; US-10-117-323-40
```

```
Query Match 17.9%; Score 99; DB 9; Length 272;
Best Local Similarity 39.6%; Pred. No. 0.0022;
Matches 19; Conservative 11; Mismatches 18; Indels 0; Gaps 0;
```

```
Qy 46 GFDIAIWMIAQWNLQSGIRVSLPQPSDIPPGTGVFIYGRDND 93
Db 116 GADVALLLEEVVNVSTHPIPLPASETFFPGMPCWVGMDVND 163
```

```
RESULT 6
US-09-923-779-152
; Sequence 152, Application US/09923779
; Patent No. US20020076721A1
; GENERAL INFORMATION:
; APPLICANT: Pyle, Ruth A.
; APPLICANT: Xu, Jiangchun
; APPLICANT: Kalos, Michael D.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
; FILE REFERENCE: 210121.553
; CURRENT APPLICATION NUMBER: US/09/923,779
```



```

; CURRENT FILING DATE: 2001-08-06
; NUMBER OF SEQ ID NOS: 155
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 152
; LENGTH: 270
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-923-779-152

Query Match      17.3%; Score 95.5; DB 10; Length 270;
Best Local Similarity 29.1%; Pred. No. 0.0053;
Matches 25; Conservative 16; Mismatches 34; Indels 11; Gaps 1

Oy      4  LTLRNGDGGIHNGSGVAVAPGVPYMSCASARQRRIAGTSGFDIAIYMLQMVNLQSG 63
      |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db      90  LAVKGPEVATINSEELVHPLMNRSCVAC-----GNDIALIKLSRAQLGDA 138

Oy      64  IRVLSLPQSPDIPPGTGVFIYGYGR 89
      ::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db     139  VQLASLPAGDILPNKTPCYITGWGR 164

RESULT 7
US-09-925-297-695
; Sequence 695, Application US/09925297
; Patent No. US20020081659A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
; FILE REFERENCE: PA105
; CURRENT APPLICATION NUMBER: US/09/925,297
; CURRENT FILING DATE: 2001-08-10
; PRIOR APPLICATION NUMBER: PCT/US00/05989
; PRIOR FILING DATE: 2000-03-08
; PRIOR APPLICATION NUMBER: 60/124,270
; PRIOR FILING DATE: 1999-03-12
; NUMBER OF SEQ ID NOS: 928
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 695
; LENGTH: 273
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (27)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (28)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (34)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-09-925-297-695

Query Match      17.3%; Score 95.5; DB 10; Length 273;
Best Local Similarity 29.1%; Pred. No. 0.0054;
Matches 25; Conservative 16; Mismatches 34; Indels 11; Gaps 1,

Oy      4  LTLRNGDGGIHNGSGVAVAPGVPYMSCASARQRRIAGTSGFDIAIYMLQMVNLQSG 63
      |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db      93  LAVKGPEVATINSEELVHPLMNRSCVAC-----GNDIALIKLSRAQLGDA 141

Oy      64  IRVLSLPQSPDIPPGTGVFIYGYGR 89
      ::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db     142  VQLASLPAGDILPNKTPCYITGWGR 167

RESULT 8
US-10-020-441-3
; Sequence 3, Application US/10020441
; Publication No. US20020182224A1
; GENERAL INFORMATION:
; APPLICANT: Doenhoff, Michael

```

[illegible]



```
; PRIOR FILING DATE: 1998-03-31
; PRIOR APPLICATION NUMBER: 60/081203
; PRIOR FILING DATE: 1998-04-09
; PRIOR APPLICATION NUMBER: 60/081229
; PRIOR FILING DATE: 1998-04-09
; PRIOR APPLICATION NUMBER: 60/081695
; PRIOR FILING DATE: 1998-04-14
; PRIOR APPLICATION NUMBER: 60/081817
; PRIOR FILING DATE: 1998-04-15
; PRIOR APPLICATION NUMBER: 60/081818
; PRIOR FILING DATE: 1998-04-15
; PRIOR APPLICATION NUMBER: 60/082999
; PRIOR FILING DATE: 1998-04-24
; PRIOR APPLICATION NUMBER: 60/083322
; PRIOR FILING DATE: 1998-04-28
; PRIOR APPLICATION NUMBER: 60/083545
; PRIOR FILING DATE: 1998-04-29
; PRIOR APPLICATION NUMBER: 60/084600
; PRIOR FILING DATE: 1998-05-07
; PRIOR APPLICATION NUMBER: 60/084627
; PRIOR FILING DATE: 1998-05-07
; PRIOR APPLICATION NUMBER: 60/084637
; PRIOR FILING DATE: 1998-05-07
; PRIOR APPLICATION NUMBER: 60/085149
; PRIOR FILING DATE: 1998-05-12
; PRIOR APPLICATION NUMBER: 60/085323
; PRIOR FILING DATE: 1998-05-13
; PRIOR APPLICATION NUMBER: 60/085338
; PRIOR FILING DATE: 1998-05-13
; PRIOR APPLICATION NUMBER: 60/085339
; PRIOR FILING DATE: 1998-05-13
; PRIOR APPLICATION NUMBER: 60/085579
; PRIOR FILING DATE: 1998-05-15
; PRIOR APPLICATION NUMBER: 60/085697
; PRIOR FILING DATE: 1998-05-15
; PRIOR APPLICATION NUMBER: 60/085704
; PRIOR FILING DATE: 1998-05-15
; PRIOR APPLICATION NUMBER: 60/086414
; PRIOR FILING DATE: 1998-05-22
; PRIOR APPLICATION NUMBER: 60/086430
; PRIOR FILING DATE: 1998-05-22
; PRIOR APPLICATION NUMBER: 60/087106
; PRIOR FILING DATE: 1998-05-28
; PRIOR APPLICATION NUMBER: 60/088026
; PRIOR FILING DATE: 1998-06-04
; PRIOR APPLICATION NUMBER: 60/088730
; PRIOR FILING DATE: 1998-06-10
; PRIOR APPLICATION NUMBER: 60/088741
; PRIOR FILING DATE: 1998-06-10
; PRIOR APPLICATION NUMBER: 60/088810
; PRIOR FILING DATE: 1998-06-10
; PRIOR APPLICATION NUMBER: 60/088858
; PRIOR FILING DATE: 1998-06-11
; PRIOR APPLICATION NUMBER: 60/089532
; PRIOR FILING DATE: 1998-06-17
; PRIOR APPLICATION NUMBER: 60/089599
; PRIOR FILING DATE: 1998-06-17
; PRIOR APPLICATION NUMBER: 60/089907
; PRIOR FILING DATE: 1998-06-18
; PRIOR APPLICATION NUMBER: 60/089947
; PRIOR FILING DATE: 1998-06-19
; PRIOR APPLICATION NUMBER: 60/090349
; PRIOR FILING DATE: 1998-06-23
; PRIOR APPLICATION NUMBER: 60/090429
; PRIOR FILING DATE: 1998-06-24
; PRIOR APPLICATION NUMBER: 60/090445
; PRIOR FILING DATE: 1998-06-24
; PRIOR APPLICATION NUMBER: 60/090538
; PRIOR FILING DATE: 1998-06-24
; PRIOR APPLICATION NUMBER: 60/090863
; PRIOR FILING DATE: 1998-06-26
; PRIOR APPLICATION NUMBER: 60/091360
; PRIOR FILING DATE: 1998-07-01
```

```
; PRIOR APPLICATION NUMBER: 60/091519
; PRIOR FILING DATE: 1998-07-02
; PRIOR APPLICATION NUMBER: 60/091982
; PRIOR FILING DATE: 1998-07-07
```

```
Query Match 15.1%; Score 83.5; DB 9; Length 290;
Best Local Similarity 32.6%; Pred. No. 0.13; 38; Indels 9; Gaps 4;
Matches 29; Conservative 13; Mismatches
```

```
Oy 7 RRGDQGGIHHQPSGVK--VAPGYMPSCSARQR---PIAQ-TLSGFDIAIVMLAQMVN 59
Db 78 RNTSETSLYQVLLGARQLVQPG--PHANYARVQVESNPLYQGTASSADVALVELAEVP 135
Oy 60 LQSGIRVLSLPQSDIPPEPGTGVFVGYG 88
Db 136 FTNYILPVCLPDPSPVIFETGMNCWVTGNG 164
```

## RESULT 12

```
US-10-121-049-222
; Sequence 222, Application US/10121049
; Publication No. US2003002239A1
```

## GENERAL INFORMATION:

```
; APPLICANT: Baker, Kevin P.
; APPLICANT: Beresini, Maureen
; APPLICANT: DeForge, Laura
; APPLICANT: Desnoyers, Luc
; APPLICANT: Filvaroff, Ellen
; APPLICANT: Gao, Wei-Qiang
; APPLICANT: Gerritsen, Mary E.
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Sherwood, Steven
; APPLICANT: Smith, Victoria
; APPLICANT: Stewart, Timothy A.
; APPLICANT: Tumas, Daniel
; APPLICANT: Watanabe, Colin K
; APPLICANT: Wood, William
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; FILE REFERENCE: P33081C17
; CURRENT APPLICATION NUMBER: US/10/121,049
; CURRENT FILING DATE: 2002-04-12
; Prior Application removed - See File Wrapper or Palm
; NUMBER OF SEQ ID NOS: 550
; SEQ ID NO 222
; LENGTH: 290
; TYPE: PRT
; ORGANISM: Homo Sapien
```

## US-10-121-049-222

```
Query Match 15.1%; Score 83.5; DB 9; Length 290;
Best Local Similarity 32.6%; Pred. No. 0.13; 38; Indels 9; Gaps 4;
Matches 29; Conservative 13; Mismatches
```

```
Oy 7 RRGDQGGIHHQPSGVK--VAPGYMPSCSARQR---PIAQ-TLSGFDIAIVMLAQMVN 59
Db 78 RNTSETSLYQVLLGARQLVQPG--PHANYARVQVESNPLYQGTASSADVALVELAEVP 135
Oy 60 LQSGIRVLSLPQSDIPPEPGTGVFVGYG 88
Db 136 FTNYILPVCLPDPSPVIFETGMNCWVTGNG 164
```

## RESULT 13

```
US-10-123-904-222
; Sequence 222, Application US/10123904
; Publication No. US20030022328A1
```

## GENERAL INFORMATION:

```
; APPLICANT: Baker, Kevin P.
; APPLICANT: Beresini, Maureen
```

```

; APPLICANT: DeForge, Laura
; APPLICANT: Desnoyers, Luc
; APPLICANT: Filvaroff, Ellen
; APPLICANT: Gao, Wei-Qiang
; APPLICANT: Gerritsen, Mary E.
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Sherwood, Steven
; APPLICANT: Smith, Victoria
; APPLICANT: Stewart, Timothy A.
; APPLICANT: Tumas, Daniel
; APPLICANT: Watanabe, Colin K
; APPLICANT: Wood, William
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; FILE REFERENCE: P3330R1C54
; CURRENT APPLICATION NUMBER: US/10/123,904
; PRIOR APPLICATION REMOVED - See File Wrapper or Palm
; NUMBER OF SEQ ID NOS: 550
; SEQ ID NO 222
; LENGTH: 290
; TYPE: PRT
; ORGANISM: Homo Sapien
; US-10-123-904-222
```

```

Query Match          15.1%; Score 83.5; DB 9; Length 290;
Best Local Similarity 32.6%; Pred. No. 0.13;
Matches 29; Conservative 13; Mismatches 38; Indels 9; Gaps 4;
```

```

Qy 7 RNGDQGIHQPSGVK--VAPGYMPCMSARQRR---PIAQ-TLSGFDIAIWLAMQVN 59
    ||:::|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 78 RNTSETSLYQVLGARQLVQPG--PHAMVAVRQVESNPLYGTASSADVALVELAEPVP 135
Qy 60 LOSGIRVLSLPQSDIPPPGCVFIVGYG 88
    :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 136 FTNYILPVCLPDPSPVIFETGMNCWVTGWG 164
```

```

RESULT 14
US-10-140-470-222
; Sequence 222, Application US/10140470
; Publication No. US2003002231A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Beresini, Maureen
; APPLICANT: DeForge, Laura
; APPLICANT: Desnoyers, Luc
; APPLICANT: Filvaroff, Ellen
; APPLICANT: Gao, Wei-Qiang
; APPLICANT: Gerritsen, Mary E.
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Sherwood, Steven
; APPLICANT: Smith, Victoria
; APPLICANT: Stewart, Timothy A.
; APPLICANT: Tumas, Daniel
; APPLICANT: Watanabe, Colin K
; APPLICANT: Wood, William
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; FILE REFERENCE: P3330R1C160
; CURRENT APPLICATION NUMBER: US/10/140,470
; PRIOR APPLICATION REMOVED - See File Wrapper or Palm
; NUMBER OF SEQ ID NOS: 550
; SEQ ID NO 222
; LENGTH: 290
; TYPE: PRT
```

```

; ORGANISM: Homo Sapien
; US-10-140-470-222
```

```

Query Match          15.1%; Score 83.5; DB 9; Length 290;
Best Local Similarity 32.6%; Pred. No. 0.13;
Matches 29; Conservative 13; Mismatches 38; Indels 9; Gaps 4;
```

```

Qy 7 RNGDQGIHQPSGVK--VAPGYMPCMSARQRR---PIAQ-TLSGFDIAIWLAMQVN 59
    ||:::|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 78 RNTSETSLYQVLGARQLVQPG--PHAMVAVRQVESNPLYGTASSADVALVELAEPVP 135
Qy 60 LOSGIRVLSLPQSDIPPPGCVFIVGYG 88
    :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 136 FTNYILPVCLPDPSPVIFETGMNCWVTGWG 164
```

```

RESULT 15
US-10-175-746-222
; Sequence 222, Application US/10175746
; Publication No. US20030027270A1
; GENERAL INFORMATION:
```

```

; APPLICANT: Baker, Kevin P.
; APPLICANT: Beresini, Maureen
; APPLICANT: DeForge, Laura
; APPLICANT: Desnoyers, Luc
; APPLICANT: Filvaroff, Ellen
; APPLICANT: Gao, Wei-Qiang
; APPLICANT: Gerritsen, Mary E.
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Sherwood, Steven
; APPLICANT: Smith, Victoria
; APPLICANT: Stewart, Timothy A.
; APPLICANT: Tumas, Daniel
; APPLICANT: Watanabe, Colin K
; APPLICANT: Wood, William
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; FILE REFERENCE: P3330R1C53
; CURRENT APPLICATION NUMBER: US/10/175,746
; PRIOR APPLICATION REMOVED - See File Wrapper or Palm
; NUMBER OF SEQ ID NOS: 550
; SEQ ID NO 222
; LENGTH: 290
; TYPE: PRT
; ORGANISM: Homo Sapien
; US-10-175-746-222
```

```

Query Match          15.1%; Score 83.5; DB 9; Length 290;
Best Local Similarity 32.6%; Pred. No. 0.13;
Matches 29; Conservative 13; Mismatches 38; Indels 9; Gaps 4;
```

```

Qy 7 RNGDQGIHQPSGVK--VAPGYMPCMSARQRR---PIAQ-TLSGFDIAIWLAMQVN 59
    ||:::|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 78 RNTSETSLYQVLGARQLVQPG--PHAMVAVRQVESNPLYGTASSADVALVELAEPVP 135
Qy 60 LOSGIRVLSLPQSDIPPPGCVFIVGYG 88
    :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 136 FTNYILPVCLPDPSPVIFETGMNCWVTGWG 164
```

```

RESULT 16
US-10-176-918-222
; Sequence 222, Application US/10176918
; Publication No. US20030027275A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Beresini, Maureen
; APPLICANT: DeForge, Laura
; APPLICANT: Desnoyers, Luc
```

```

; APPLICANT: Filvaroff, Ellen
; APPLICANT: Gao, Wei-Qiang
; APPLICANT: Gerritsen, Mary E.
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Sherwood, Steven
; APPLICANT: Smith, Victoria
; APPLICANT: Stewart, Timothy A.
; APPLICANT: Tumas, Daniel
; APPLICANT: Watanabe, Colin K
; APPLICANT: Wood, William
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; TITLE OF INVENTION: ACIDS ENCODING THE SAME
; FILE REFERENCE: P3330R1C382
; CURRENT APPLICATION NUMBER: US/10/176,918
; CURRENT FILING DATE: 2002-06-20
; Prior Application removed - See File Wrapper or Palm
; NUMBER OF SEQ ID NOS: 550
; SEQ ID NO 222
; LENGTH: 290
; TYPE: PRT
; ORGANISM: Homo Sapien
US-10-176-918-222
```

```

Query Match      15.1%; Score 83.5; DB 9; Length 290;
Best Local Similarity 32.6%; Pred. No. 0.13;
Matches 29; Conservative 13; Mismatches 38; Indels 9; Gaps 4;
```

```

Qy 7 RNGDQGGIHQPSGVK--VAPGYMPCMSARQR---PIAQ-TLSGFDIAIWLAAQWVN 59
    || : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 78 RNTSETSLYQLVIGARQLVQPG--PHANYARVQVESNPLYQGTASSADVALVELAEPV 135
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 60 LOSGIRVLSLPQSDIPPPGTVFIVGYG 88
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 136 FTNYILPVCLPDPSPVIFETGMNCWVTGWG 164
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
```

## RESULT 17

```

US-10-176-921-222
; Sequence 222, Application US/10176921
; Publication No. US20030027276a1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Beresini, Maureen
; APPLICANT: Deforge, Laura
; APPLICANT: Desnoyers, Luc
; APPLICANT: Filvaroff, Ellen
; APPLICANT: Gao, Wei-Qiang
; APPLICANT: Gerritsen, Mary E.
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Sherwood, Steven
; APPLICANT: Smith, Victoria
; APPLICANT: Stewart, Timothy A.
; APPLICANT: Tumas, Daniel
; APPLICANT: Watanabe, Colin K
; APPLICANT: Wood, William
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; TITLE OF INVENTION: ACIDS ENCODING THE SAME
; FILE REFERENCE: P3330R1C388
; CURRENT APPLICATION NUMBER: US/10/176,921
; CURRENT FILING DATE: 2002-06-20
; Prior Application removed - See File Wrapper or Palm
; NUMBER OF SEQ ID NOS: 550
; SEQ ID NO 222
; LENGTH: 290
; TYPE: PRT
; ORGANISM: Homo Sapien
US-10-176-921-222
```

```

Query Match      15.1%; Score 83.5; DB 9; Length 290;
Best Local Similarity 32.6%; Pred. No. 0.13;
Matches 29; Conservative 13; Mismatches 38; Indels 9; Gaps 4;
```

```

Qy 7 RNGDQGGIHQPSGVK--VAPGYMPCMSARQR---PIAQ-TLSGFDIAIWLAAQWVN 59
    || : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 78 RNTSETSLYQLVIGARQLVQPG--PHANYARVQVESNPLYQGTASSADVALVELAEPV 135
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 60 LOSGIRVLSLPQSDIPPPGTVFIVGYG 88
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 136 FTNYILPVCLPDPSPVIFETGMNCWVTGWG 164
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
```

## RESULT 18

```

US-10-137-865-222
; Sequence 222, Application US/10137865
; Publication No. US20030032155a1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Beresini, Maureen
; APPLICANT: Deforge, Laura
; APPLICANT: Desnoyers, Luc
; APPLICANT: Filvaroff, Ellen
; APPLICANT: Gao, Wei-Qiang
; APPLICANT: Gerritsen, Mary E.
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Sherwood, Steven
; APPLICANT: Smith, Victoria
; APPLICANT: Stewart, Timothy A.
; APPLICANT: Tumas, Daniel
; APPLICANT: Watanabe, Colin K
; APPLICANT: Wood, William
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; TITLE OF INVENTION: ACIDS ENCODING THE SAME
; FILE REFERENCE: P3330R1C154
; CURRENT APPLICATION NUMBER: US/10/137,865
; CURRENT FILING DATE: 2002-05-03
; Prior Application removed - See Palm or File Wrapper
; NUMBER OF SEQ ID NOS: 550
; SEQ ID NO 222
; LENGTH: 290
; TYPE: PRT
; ORGANISM: Homo Sapien
US-10-137-865-222
```

```

Query Match      15.1%; Score 83.5; DB 9; Length 290;
Best Local Similarity 32.6%; Pred. No. 0.13;
Matches 29; Conservative 13; Mismatches 38; Indels 9; Gaps 4;
```

```

Qy 7 RNGDQGGIHQPSGVK--VAPGYMPCMSARQR---PIAQ-TLSGFDIAIWLAAQWVN 59
    || : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 78 RNTSETSLYQLVIGARQLVQPG--PHANYARVQVESNPLYQGTASSADVALVELAEPV 135
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 60 LOSGIRVLSLPQSDIPPPGTVFIVGYG 88
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 136 FTNYILPVCLPDPSPVIFETGMNCWVTGWG 164
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
```

## RESULT 19

```

US-10-140-474-222
; Sequence 222, Application US/10140474
; Publication No. US20030032156a1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Beresini, Maureen
; APPLICANT: Deforge, Laura
; APPLICANT: Desnoyers, Luc
; APPLICANT: Filvaroff, Ellen
; APPLICANT: Gao, Wei-Qiang
US-10-140-474-222
```

```

; APPLICANT: Gerritsen, Mary E.
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Sherwood, Steven
; APPLICANT: Smith, Victoria
; APPLICANT: Stewart, Timothy A.
; APPLICANT: Tumas, Daniel
; APPLICANT: Watanabe, Colin K
; APPLICANT: Wood, William
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; FILE REFERENCE: P330R1C162
; CURRENT APPLICATION NUMBER: US/10/140,474
; CURRENT FILING DATE: 2002-05-06
; Prior Application removed - See Palm or File Wrapper
; NUMBER OF SEQ ID NOS: 550
; SEQ ID NO 222
; LENGTH: 290
; TYPE: PRT
; ORGANISM: Homo Sapien
US-10-140-474-222
```

```

Query Match          15.1%; Score 83.5; DB 9; Length 290;
Best Local Similarity 32.6%; Pred. No. 0.13;
Matches 29; Conservative 13; Mismatches 38; Indels 9; Gaps 4;
```

```

Qy 7 RRGDQGIHQHSGVK--VAPGYMPCMSARQR---PIAQ-TLSGPDIAVMLAQMN 59
    | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 78 RNTSETSLYQVLLGARQLVQPG--PHANYARVQVESNPLYQGTASSADVALVELEAVP 135
    : | : | : | : | : | : | : | : | : | : | : | : |
Qy 60 LQSGIRVISLPQPSDIPPGTGVFIVGYG 88
    : | : | : | : | : | : | : | : | : | : | : | : |
Db 136 FTNYILPVCLPDPSPVIFETGMNCWVTGWG 164
    : | : | : | : | : | : | : | : | : | : | : | : |
```

```

RESULT 20
US-10-142-431-222
; Sequence 222, Application US/10142431
; Publication No. US20030036179A1
; GENERAL INFORMATION:
```

```

; APPLICANT: Baker, Kevin P.
; APPLICANT: Beresini, Maureen
; APPLICANT: DeForge, Laura
; APPLICANT: Desnoyers, Luc
; APPLICANT: Filvaroff, Ellen
; APPLICANT: Gao, Wei-Qiang
; APPLICANT: Gerritsen, Mary E.
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Sherwood, Steven
; APPLICANT: Smith, Victoria
; APPLICANT: Stewart, Timothy A.
; APPLICANT: Tumas, Daniel
; APPLICANT: Watanabe, Colin K
; APPLICANT: Wood, William
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; FILE REFERENCE: P330R1C251
; CURRENT APPLICATION NUMBER: US/10/142,431
; CURRENT FILING DATE: 2002-05-10
; Prior Application removed - See File Wrapper or Palm
; NUMBER OF SEQ ID NOS: 550
; SEQ ID NO 222
; LENGTH: 290
; TYPE: PRT
; ORGANISM: Homo Sapien
US-10-142-431-222
```

```

Query Match          15.1%; Score 83.5; DB 9; Length 290;
```

```

Best Local Similarity 32.6%; Pred. No. 0.13;
Matches 29; Conservative 13; Mismatches 38; Indels 9; Gaps 4;

Qy 7 RRGDQGIHQHSGVK--VAPGYMPCMSARQR---PIAQ-TLSGPDIAVMLAQMN 59
    | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 78 RNTSETSLYQVLLGARQLVQPG--PHANYARVQVESNPLYQGTASSADVALVELEAVP 135
    : | : | : | : | : | : | : | : | : | : | : | : |
Qy 60 LQSGIRVISLPQPSDIPPGTGVFIVGYG 88
    : | : | : | : | : | : | : | : | : | : | : | : |
Db 136 FTNYILPVCLPDPSPVIFETGMNCWVTGWG 164
    : | : | : | : | : | : | : | : | : | : | : | : |
```

```

RESULT 21
US-10-143-114-222
; Sequence 222, Application US/10143114
; Publication No. US20030036180A1
; GENERAL INFORMATION:
```

```

; APPLICANT: Baker, Kevin P.
; APPLICANT: Beresini, Maureen
; APPLICANT: DeForge, Laura
; APPLICANT: Desnoyers, Luc
; APPLICANT: Filvaroff, Ellen
; APPLICANT: Gao, Wei-Qiang
; APPLICANT: Gerritsen, Mary E.
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Sherwood, Steven
; APPLICANT: Smith, Victoria
; APPLICANT: Stewart, Timothy A.
; APPLICANT: Tumas, Daniel
; APPLICANT: Watanabe, Colin K
; APPLICANT: Wood, William
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; FILE REFERENCE: P330R1C211
; CURRENT APPLICATION NUMBER: US/10/143,114
; CURRENT FILING DATE: 2002-05-09
; Prior Application removed - See Palm or File Wrapper
; NUMBER OF SEQ ID NOS: 550
; SEQ ID NO 222
; LENGTH: 290
; TYPE: PRT
; ORGANISM: Homo Sapien
US-10-143-114-222
```

```

Query Match          15.1%; Score 83.5; DB 9; Length 290;
Best Local Similarity 32.6%; Pred. No. 0.13;
Matches 29; Conservative 13; Mismatches 38; Indels 9; Gaps 4;
```

```

Qy 7 RRGDQGIHQHSGVK--VAPGYMPCMSARQR---PIAQ-TLSGPDIAVMLAQMN 59
    | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 78 RNTSETSLYQVLLGARQLVQPG--PHANYARVQVESNPLYQGTASSADVALVELEAVP 135
    : | : | : | : | : | : | : | : | : | : | : | : |
Qy 60 LQSGIRVISLPQPSDIPPGTGVFIVGYG 88
    : | : | : | : | : | : | : | : | : | : | : | : |
Db 136 FTNYILPVCLPDPSPVIFETGMNCWVTGWG 164
    : | : | : | : | : | : | : | : | : | : | : | : |
```

```

RESULT 22
US-10-140-002-222
; Sequence 222, Application US/10140002
; Publication No. US20030037623A1
; GENERAL INFORMATION:
```

```

; APPLICANT: Baker, Kevin P.
; APPLICANT: Beresini, Maureen
; APPLICANT: DeForge, Laura
; APPLICANT: Desnoyers, Luc
; APPLICANT: Filvaroff, Ellen
; APPLICANT: Gao, Wei-Qiang
; APPLICANT: Gerritsen, Mary E.
; APPLICANT: Goddard, Audrey
```

```

; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Sherwood, Steven
; APPLICANT: Smith, Victoria
; APPLICANT: Stewart, Timothy A.
; APPLICANT: Tumas, Daniel
; APPLICANT: Watanabe, Colin K
; APPLICANT: Wood, William
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; FILE REFERENCE: P3330R1C59
; CURRENT FILING DATE: 2002-05-06
; Prior Application removed - See Palm or File Wrapper
; NUMBER OF SEQ ID NOS: 550
; SEQ ID NO 222
; LENGTH: 290
; TYPE: PRT
; ORGANISM: Homo Sapien
US-10-140-002-222

Query Match          15.1%; Score 83.5; DB 9; Length 290;
Best Local Similarity 32.6%; Pred. No. 0.13;
Matches 29; Conservative 13; Mismatches 38; Indels 9; Gaps 4;

OY 7 RRGDQGGIHQPSGVK--VAPGYMPCMSARQR---PIAQ-TLSGFDIAIYMLAQMVN 59
DB 78 RNTSETSLYQVLLGARQLVQPG--PHANYARVQVESNPLVGTASADVALVELAPVP 135
OY 60 LOSGIRVLSLPQSDIPPPGTGVFIVGYG 88
DB 136 FNNYILPVCLPDPSPVIFETGNMCWVTGWG 164

RESULT 23
US-10-142-419-222
; Sequence 222, Application US/10142419
; Publication No. US2003004945A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Beresini, Maureen
; APPLICANT: DeForge, Laura
; APPLICANT: Desnoyers, Luc
; APPLICANT: Filvaroff, Ellen
; APPLICANT: Gao, Wei-Qiang
; APPLICANT: Gerritsen, Mary E.
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Sherwood, Steven
; APPLICANT: Smith, Victoria
; APPLICANT: Stewart, Timothy A.
; APPLICANT: Tumas, Daniel
; APPLICANT: Watanabe, Colin K
; APPLICANT: Wood, William
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; FILE REFERENCE: P3330R1C244
; CURRENT FILING DATE: 2002-05-10
; Prior Application removed - See File Wrapper or Palm
; NUMBER OF SEQ ID NOS: 550
; SEQ ID NO 222
; LENGTH: 290
; TYPE: PRT
; ORGANISM: Homo Sapien
US-10-142-419-222

Query Match          15.1%; Score 83.5; DB 9; Length 290;
Best Local Similarity 32.6%; Pred. No. 0.13;
Matches 29; Conservative 13; Mismatches 38; Indels 9; Gaps 4;
```

```

OY 7 RRGDQGGIHQPSGVK--VAPGYMPCMSARQR---PIAQ-TLSGFDIAIYMLAQMVN 59
DB 78 RNTSETSLYQVLLGARQLVQPG--PHANYARVQVESNPLVGTASADVALVELAPVP 135
OY 60 LOSGIRVLSLPQSDIPPPGTGVFIVGYG 88
DB 136 FNNYILPVCLPDPSPVIFETGNMCWVTGWG 164

RESULT 24
US-10-123-262-222
; Sequence 222, Application US/10123262
; Publication No. US20030049816A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Beresini, Maureen
; APPLICANT: DeForge, Laura
; APPLICANT: Desnoyers, Luc
; APPLICANT: Filvaroff, Ellen
; APPLICANT: Gao, Wei-Qiang
; APPLICANT: Gerritsen, Mary E.
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Sherwood, Steven
; APPLICANT: Smith, Victoria
; APPLICANT: Stewart, Timothy A.
; APPLICANT: Tumas, Daniel
; APPLICANT: Watanabe, Colin K
; APPLICANT: Wood, William
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; FILE REFERENCE: P3330R1C38
; CURRENT FILING DATE: 2002-04-15
; Prior Application removed - See File Wrapper or Palm
; NUMBER OF SEQ ID NOS: 550
; SEQ ID NO 222
; LENGTH: 290
; TYPE: PRT
; ORGANISM: Homo Sapien
US-10-123-262-222

Query Match          15.1%; Score 83.5; DB 9; Length 290;
Best Local Similarity 32.6%; Pred. No. 0.13;
Matches 29; Conservative 13; Mismatches 38; Indels 9; Gaps 4;

OY 7 RRGDQGGIHQPSGVK--VAPGYMPCMSARQR---PIAQ-TLSGFDIAIYMLAQMVN 59
DB 78 RNTSETSLYQVLLGARQLVQPG--PHANYARVQVESNPLVGTASADVALVELAPVP 135
OY 60 LOSGIRVLSLPQSDIPPPGTGVFIVGYG 88
DB 136 FNNYILPVCLPDPSPVIFETGNMCWVTGWG 164

RESULT 25
US-10-142-423-222
; Sequence 222, Application US/10142423
; Publication No. US20030049817A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Beresini, Maureen
; APPLICANT: DeForge, Laura
; APPLICANT: Desnoyers, Luc
; APPLICANT: Filvaroff, Ellen
; APPLICANT: Gao, Wei-Qiang
; APPLICANT: Gerritsen, Mary E.
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
```

```
; APPLICANT: Sherwood, Steven
; APPLICANT: Smith, Victoria
; APPLICANT: Stewart, Timothy A.
; APPLICANT: Tumas, Daniel
; APPLICANT: Watanabe, Colin K
; APPLICANT: Wood, William
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; FILE REFERENCE: P330R1C249
; CURRENT APPLICATION NUMBER: US/10/142,423
; CURRENT FILING DATE: 2002-05-10
; Prior Application removed - See File Wrapper or Palm
; NUMBER OF SEQ ID NOS: 550
; SEQ ID NO 222
; LENGTH: 290
; TYPE: PRT
; ORGANISM: Homo Sapien
US-10-142-423-222
```

Query Match 15.1%; Score 83.5; DB 9; Length 290;

Best Local Similarity 32.6%; Pred. No. 0.13; Mismatches 38; Indels 9; Gaps 4;

Matches 29; Conservative 13; Mismatches 38; Indels 9; Gaps 4;

```
Qy 7 RNDGQGIHOPSGVK--VAPGYMPCMSARQRR---PIAQ-TLISGFDIAIWMIAQWVN 59
|| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 78 RNTSETSLYQVLGARQLVOPG--PHAMYARVQVESNPLYQGTASSADVALVELAEPV 135
: : : : : : : : : : : : : : : : : : : : : : : : : : : :
```

Qy 60 LOSGIRVLSLPQPSDIPPPGTGVFIYGYG 88

Db 136 FTNYILPVCLPDPSPVIFETGMNCWVTGWG 164

RESULT 26

US-10-121-050-222

; Sequence 222, Application US/10121050

; Publication No. US20030054516A1

; GENERAL INFORMATION:

; APPLICANT: Baker, Kevin P.

; APPLICANT: Beresini, Maureen

; APPLICANT: Deforge, Laura

; APPLICANT: Desnoyers, Luc

; APPLICANT: Filvaroff, Ellen

; APPLICANT: Gao, Wei-Qiang

; APPLICANT: Gerritsen, Mary E.

; APPLICANT: Goddard, Audrey

; APPLICANT: Godowski, Paul J.

; APPLICANT: Gurney, Austin L.

; APPLICANT: Sherwood, Steven

; APPLICANT: Smith, Victoria

; APPLICANT: Stewart, Timothy A.

; APPLICANT: Tumas, Daniel

; APPLICANT: Watanabe, Colin K

; APPLICANT: Wood, William

; APPLICANT: Zhang, Zemin

; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC

; FILE REFERENCE: P330R1C20

; CURRENT APPLICATION NUMBER: US/10/121,050

; CURRENT FILING DATE: 2002-04-12

; Prior Application removed - See File Wrapper or Palm

; NUMBER OF SEQ ID NOS: 550

; SEQ ID NO 222

; LENGTH: 290

; TYPE: PRT

; ORGANISM: Homo Sapien

US-10-121-050-222

Query Match 15.1%; Score 83.5; DB 9; Length 290;

Best Local Similarity 32.6%; Pred. No. 0.13; Mismatches 38; Indels 9; Gaps 4;

Matches 29; Conservative 13; Mismatches 38; Indels 9; Gaps 4;

Qy 7 RNDGQGIHOPSGVK--VAPGYMPCMSARQRR---PIAQ-TLISGFDIAIWMIAQWVN 59

```
Db 78 RNTSETSLYQVLGARQLVOPG--PHAMYARVQVESNPLYQGTASSADVALVELAEPV 135
|| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 60 LOSGIRVLSLPQPSDIPPPGTGVFIYGYG 88
: : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 136 FTNYILPVCLPDPSPVIFETGMNCWVTGWG 164
: : : : : : : : : : : : : : : : : : : : : : : : : : : :
```

RESULT 27

US-10-141-755-222

; Sequence 222, Application US/10141755

; Publication No. US20030054517A1

; GENERAL INFORMATION:

; APPLICANT: Baker, Kevin P.

; APPLICANT: Beresini, Maureen

; APPLICANT: Deforge, Laura

; APPLICANT: Desnoyers, Luc

; APPLICANT: Filvaroff, Ellen

; APPLICANT: Gao, Wei-Qiang

; APPLICANT: Gerritsen, Mary E.

; APPLICANT: Goddard, Audrey

; APPLICANT: Godowski, Paul J.

; APPLICANT: Gurney, Austin L.

; APPLICANT: Sherwood, Steven

; APPLICANT: Smith, Victoria

; APPLICANT: Stewart, Timothy A.

; APPLICANT: Tumas, Daniel

; APPLICANT: Watanabe, Colin K

; APPLICANT: Wood, William

; APPLICANT: Zhang, Zemin

; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC

; FILE REFERENCE: P330R1C192

; CURRENT APPLICATION NUMBER: US/10/141,755

; CURRENT FILING DATE: 2002-05-08

; Prior Application removed - See File Wrapper or Palm

; NUMBER OF SEQ ID NOS: 550

; SEQ ID NO 222

; LENGTH: 290

; TYPE: PRT

; ORGANISM: Homo Sapien

US-10-141-755-222

Query Match 15.1%; Score 83.5; DB 9; Length 290;

Best Local Similarity 32.6%; Pred. No. 0.13; Mismatches 38; Indels 9; Gaps 4;

Matches 29; Conservative 13; Mismatches 38; Indels 9; Gaps 4;

```
Qy 7 RNDGQGIHOPSGVK--VAPGYMPCMSARQRR---PIAQ-TLISGFDIAIWMIAQWVN 59
|| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 78 RNTSETSLYQVLGARQLVOPG--PHAMYARVQVESNPLYQGTASSADVALVELAEPV 135
: : : : : : : : : : : : : : : : : : : : : : : : : : : :
```

Qy 60 LOSGIRVLSLPQPSDIPPPGTGVFIYGYG 88

Db 136 FTNYILPVCLPDPSPVIFETGMNCWVTGWG 164

RESULT 28

US-10-143-032-222

; Sequence 222, Application US/10143032

; Publication No. US2003005909A1

; GENERAL INFORMATION:

; APPLICANT: Baker, Kevin P.

; APPLICANT: Beresini, Maureen

; APPLICANT: Deforge, Laura

; APPLICANT: Desnoyers, Luc

; APPLICANT: Filvaroff, Ellen

; APPLICANT: Gao, Wei-Qiang

; APPLICANT: Gerritsen, Mary E.

; APPLICANT: Goddard, Audrey

; APPLICANT: Godowski, Paul J.

; APPLICANT: Gurney, Austin L.

; APPLICANT: Sherwood, Steven

; APPLICANT: Smith, Victoria







```

; CURRENT APPLICATION NUMBER: US/10/041,400A
; CURRENT FILING DATE: 2002-01-08
; PRIOR APPLICATION NUMBER: US/09/387,375
; PRIOR FILING DATE: 1999-08-31
; NUMBER OF SEQ ID NOS: 9
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 7
; LENGTH: 284
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-041-400A-7

Query Match      14.5%; Score 80; DB 12; Length 284;
Best Local Similarity 36.6%; Pred. No. 0.3;
Matches 15; Conservative 10; Mismatches 16; Indels 0; Gaps 0;

Qy      48 DIAIVMLAQMWNLOSIGIRVLSLPPPSDIPPPGTVGVIVGYG 88
Db      126 DLALQLRRPVPPLSARVQPVCLPVGARPPPGTPCRVTGWG 166

RESULT 37
US-10-041-264A-7
; Sequence 7, Application US/10041264A
; Patent No. US20020142446A1
; GENERAL INFORMATION:
; APPLICANT: Darrow, Andrew
; APPLICANT: Andrade-Gordon, Patricia
; APPLICANT: Qi, Jensen
; TITLE OF INVENTION: DNA Encoding the Human Serine
; TITLE OF INVENTION: Protease EOS
; FILE REFERENCE: ORT-1031
; CURRENT APPLICATION NUMBER: US/10/041,264A
; CURRENT FILING DATE: 2002-01-08
; PRIOR APPLICATION NUMBER: US/09/387,375
; PRIOR FILING DATE: 1999-08-31
; NUMBER OF SEQ ID NOS: 9
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 7
; LENGTH: 284
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-041-264A-7

Query Match      14.5%; Score 80; DB 12; Length 284;
Best Local Similarity 36.6%; Pred. No. 0.3;
Matches 15; Conservative 10; Mismatches 16; Indels 0; Gaps 0;

Qy      48 DIAIVMLAQMWNLOSIGIRVLSLPPPSDIPPPGTVGVIVGYG 88
Db      126 DLALQLRRPVPPLSARVQPVCLPVGARPPPGTPCRVTGWG 166

RESULT 38
US-10-042-091A-7
; Sequence 7, Application US/10042091A
; Patent No. US20020142447A1
; GENERAL INFORMATION:
; APPLICANT: Darrow, Andrew
; APPLICANT: Andrade-Gordon, Patricia
; APPLICANT: Qi, Jensen
; TITLE OF INVENTION: DNA Encoding the Human Serine
; TITLE OF INVENTION: Protease EOS
; FILE REFERENCE: ORT-1031
; CURRENT APPLICATION NUMBER: US/10/042,091A
; CURRENT FILING DATE: 2002-01-08
; PRIOR APPLICATION NUMBER: US/09/387,375
; PRIOR FILING DATE: 1999-08-31
; NUMBER OF SEQ ID NOS: 9
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 7
; LENGTH: 284
; TYPE: PRT
```

```

; ORGANISM: Homo sapiens
US-10-042-091A-7

Query Match      14.5%; Score 80; DB 12; Length 284;
Best Local Similarity 36.6%; Pred. No. 0.3;
Matches 15; Conservative 10; Mismatches 16; Indels 0; Gaps 0;

Qy      48 DIAIVMLAQMWNLOSIGIRVLSLPPPSDIPPPGTVGVIVGYG 88
Db      126 DLALQLRRPVPPLSARVQPVCLPVGARPPPGTPCRVTGWG 166

RESULT 39
US-10-041-400A-9
; Sequence 9, Application US/10041400A
; Patent No. US20020110895A1
; GENERAL INFORMATION:
; APPLICANT: Darrow, Andrew
; APPLICANT: Andrade-Gordon, Patricia
; APPLICANT: Qi, Jensen
; TITLE OF INVENTION: DNA Encoding the Human Serine
; TITLE OF INVENTION: Protease EOS
; FILE REFERENCE: ORT-1031
; CURRENT APPLICATION NUMBER: US/10/041,400A
; CURRENT FILING DATE: 2002-01-08
; PRIOR APPLICATION NUMBER: US/09/387,375
; PRIOR FILING DATE: 1999-08-31
; NUMBER OF SEQ ID NOS: 9
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 9
; LENGTH: 316
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Amino acid
US-10-041-400A-9

Query Match      14.5%; Score 80; DB 12; Length 316;
Best Local Similarity 36.6%; Pred. No. 0.35;
Matches 15; Conservative 10; Mismatches 16; Indels 0; Gaps 0;

Qy      48 DIAIVMLAQMWNLOSIGIRVLSLPPPSDIPPPGTVGVIVGYG 88
Db      141 DLALQLRRPVPPLSARVQPVCLPVGARPPPGTPCRVTGWG 181

RESULT 40
US-10-041-264A-9
; Sequence 9, Application US/10041264A
; Patent No. US20020142446A1
; GENERAL INFORMATION:
; APPLICANT: Darrow, Andrew
; APPLICANT: Andrade-Gordon, Patricia
; APPLICANT: Qi, Jensen
; TITLE OF INVENTION: DNA Encoding the Human Serine
; TITLE OF INVENTION: Protease EOS
; FILE REFERENCE: ORT-1031
; CURRENT APPLICATION NUMBER: US/10/041,264A
; CURRENT FILING DATE: 2002-01-08
; PRIOR APPLICATION NUMBER: US/09/387,375
; PRIOR FILING DATE: 1999-08-31
; NUMBER OF SEQ ID NOS: 9
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 9
; LENGTH: 316
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Amino acid
US-10-041-264A-9
```

Query Match 14.5%; Score 80; DB 12; Length 316;  
 Best Local Similarity 36.6%; Pred. No. 0.35;  
 Matches 15; Conservative 10; Mismatches 16; Indels 0; Gaps 0;

QY 48 DIAIVMLAQMVLQSGIRVISLPQPSDIPPRGTGVFIYGYG 88  
 |::| : | : : : | | | | : : |  
 Db 141 DLALQLRRPVPLSARVQPVCLFVPGARRPPGTGTCRVTGWG 181

Search completed: April 9, 2003, 13:10:13  
 Job time : 42.0984 secs

GenCore version 5.1.4.p5.4578  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 9, 2003, 12:40:41 ; Search time 25.1967 Seconds  
(without alignments)  
174.487 Million cell updates/sec

Title: US-10-020-441-2  
Perfect score: 552  
Sequence: 1 VSFPLTNNGQCGIHHQPSG.....YGRDDNDPSPSKNGILKK 106

Scoring table: BIOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : SwissProt\_40.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	552	100.0	264	1	CERC_SCHMA
2	102	18.5	276	1	MCT6_MOUSE
3	101	18.3	270	1	TRYT_MERUN
4	100	18.1	275	1	TRYA_HUMAN
5	99	17.9	273	1	MCT7_RAT
6	99	17.9	275	1	TRB1_HUMAN
7	98	17.8	235	1	TRYD_HUMAN
8	95.5	17.3	270	1	ELJ3A_HUMAN
9	93	16.8	246	1	MCT9_MOUSE
10	93	16.8	246	1	MCTX_MOUSE
11	93	16.8	275	1	TRB2_HUMAN
12	92.5	16.8	270	1	ELJ3B_HUMAN
13	89	16.1	274	1	MCT6_RAT
14	88.5	16.0	244	1	MCT2_MOUSE
15	88.5	16.0	273	1	TRYT_SHEEP
16	88.5	16.0	275	1	TRYT_PIG
17	86	15.6	266	1	ELI_BOVIN
18	85.5	15.5	269	1	TRYM_CANFA
19	85	15.4	266	1	ELI_RAT
20	83.5	15.1	290	1	MPN_HUMAN
21	83	15.0	273	1	MCT7_MOUSE
22	82.5	14.9	855	1	STI4_HUMAN
23	81.5	14.8	247	1	MCT2_RAT
24	81	14.7	246	1	MCT4_RAT
25	81	14.7	1113	1	CORI_MOUSE
26	80	14.5	711	1	HGFL_HUMAN
27	78.5	14.2	258	1	ELI_HUMAN
28	78	14.1	321	1	TRYG_HUMAN
29	77.5	14.0	253	1	CAC3_BOVIN
30	77.5	14.0	271	1	EL2_RAT
31	76	13.8	268	1	CLCR_RAT
32	76	13.8	716	1	HGFL_MOUSE
33	76	13.8	855	1	STI4_MOUSE

34	75	13.6	260	1	GRAA_MOUSE
35	75	13.6	268	1	CLCR_HUMAN
36	74.5	13.5	246	1	MCT1_MERUN
37	74.5	13.5	246	1	MCT1_MOUSE
38	74.5	13.5	246	1	ELI_RAT
39	74	13.4	246	1	TRY2_MOUSE
40	74	13.4	256	1	PRN3_HUMAN
41	74	13.4	418	1	HAT7_HUMAN
42	73.5	13.3	271	1	EL2_MOUSE
43	73	13.2	1042	1	CORI_HUMAN
44	72	13.0	269	1	ELJ2B_HUMAN
45	72	13.0	271	1	CTRL_PENVA

## ALIGNMENTS

```
RESULT 1
CERC_SCHMA STANDARD; PRT; 264 AA.
ID AC P12546;
DT 01-OCT-1989 (Rel. 12, Last sequence update)
DT 01-OCT-1989 (Rel. 12, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Cercarial protease precursor (EC 3.4.21.-) (Cercarial elastase).
OS Schistosoma mansoni (Blood fluke).
OC Eukaryota; Metazoa; Platyhelminthes; Trematoda; Digenea; Strigoidida;
OC Schistosomatidae; Schistosomatidae; Schistosoma.
OK NCBI_TaxID=6183;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=88330818; PubMed=3166457;
RA Newport G.R., McKerrow J.H., Hedstrom R., Pettit M., McGarrigle L.,
RA Barr P.J., Agabian N.;
RT "Cloning of the proteinase that facilitates infection by schistosome
RT parasites.";
RL J. Biol. Chem. 263:13179-13184(1988).
CC -1- FUNCTION: THIS PROTEASE CLEAVES ELASTIN AND THUS FACILITATES
CC -1- PENETRATION OF SCHISTOSOME PARASITE LARVAE THROUGH ELASTIN-
CC RICH TISSUE OF THE HOST.
CC -1- ENZYME REGULATION: ACTIVATED BY AN AUTOCATALYTIC MECHANISM.
CC -1- SUBCELLULAR LOCATION: ACETABULAR (PENETRATION) GLANDS.
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; J03946; AAA29864.1; -.
DR PIR; A28942; A28942.
DR HSSP; P20231; IAAO.
DR MEROPS; S01.144; -.
DR InterPro; IPR001314; Chymotrypsin.
DR InterPro; IPR001254; Ser_protease_Try.
DR Pfam; PF00089; Trypsin; 1.
DR PRINTS; PR00722; CHYMOTRYPSIN.
DR SMART; SM00020; TRYPSIN_SPC; 1.
DR PROSITE; PS50240; TRYPSIN_DOM; 1.
DR PROSITE; PS00134; TRYPSIN_HIS; 1.
DR PROSITE; PS00135; TRYPSIN_SER; FALSE NEG.
KW Hydroxylase; Serine protease; Zymogen; Signal.
FT SIGNAL 1
FT PROPEP 20 27
FT CHAIN 28 264
FT ACT_SITE 68 68
FT ACT_SITE 126 126
FT ACT_SITE 218 218
FT SEQUENCE 264 AA; 28545 MW; E25129A7C5D5010 CRC64;
```

Query Match 100.0%; Score 552; DB 1; Length 264;  
Best Local Similarity 100.0%; Pred. No. 1,7e-53;  
Matches 106; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VSFLRLRNGDDQGIHOPGVAVAPGYMPCMSARORRPIAGTISGFDIAIWMIAQMWL 60  
DB 79 VSFLRLRNGDDQGIHOPGVAVAPGYMPCMSARORRPIAGTISGFDIAIWMIAQMWL 138

QY 61 QSGIRVLSLPQPSDIPPGTGVFIYGYGRDNDPDRSRNGGILKK 106  
DB 139 QSGIRVLSLPQPSDIPPGTGVFIYGYGRDNDPDRSRNGGILKK 184

RESULT 2  
MCT6\_MOUSE STANDARD; PRT; 276 AA.

AC P21845; 061962; (Rel. 18, Created)  
DT 01-MAY-1991 (Rel. 19, Last sequence update)  
DT 01-AUG-1991 (Rel. 19, Last annotation update)  
DT 15-JUN-2002 (Rel. 41, Last annotation update)  
DE Mast cell protease 6 precursor (EC 3.4.21.59) (MCP-6) (Trypsin).  
GN MCP6.  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
OX NCBI\_TaxID=10090;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=91139682; PubMed=1995638;  
RA Reynolds D.S., Gurley D.S., Austen K.F., Serafin W.E.;  
RT "Cloning of the cDNA and gene of mouse mast cell protease-6.  
RT Transcription by progenitor mast cells and mast cells of the  
RT connective tissue subclass.";  
RL J. Biol. Chem. 266:3847-3853(1991).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=Leaden X A1;  
RX MEDLINE=94023807; PubMed=8210998;  
RA Huang R., Abirink M., Gobl A.E., Nilsson G., Aveskog M., Larsson L.G.,  
RA Nilsson K., Hellman L.;  
RT "Expression of a mast cell tryptase in the human monocytic cell lines  
RT U-937 and Mono Mac 6.";  
RL Scand. J. Immunol. 38:359-367(1993).  
RN [3]  
RP SEQUENCE FROM N.A. (SHORT ISOFORM).  
RC STRAIN=Leaden X A1;  
RX MEDLINE=95048582; PubMed=7959952;  
RA Huang R., Hellman L.T.;  
RT "Genes for mast-cell serine protease and their molecular evolution.";  
RL Immunogenetics 40:397-414(1994).  
RN [4]  
RP SEQUENCE OF 32-54.  
RX MEDLINE=90222202; PubMed=2326280;  
RA Reynolds D.S., Stevens R.L., Lane W.S., Carr M.H., Austen K.F.,  
RA Serafin W.E.;  
RT "Different mouse mast cell populations express various combinations  
RT of at least six distinct mast cell serine proteases.";  
RL Proc. Natl. Acad. Sci. U.S.A. 87:3230-3234(1990).  
CC -1- FUNCTION: TRYPTASE IS THE MAJOR NEUTRAL PROTEASE PRESENT IN MAST  
CC CELLS, AND IS SECRETED UPON THE COUPLED ACTIVATION-DEGRADATION  
CC RESPONSE OF THIS CELL TYPE.  
CC -1- CATALYTIC ACTIVITY: Preferential cleavage: Arg-|-, Lys-|-, but  
CC with more restricted specificity than trypsin.  
CC -1- ALTERNATIVE PRODUCTS: 2 isoforms: a long form (shown here) and a  
CC short form; are produced by alternative splicing. The short form  
CC is probably a non functional variant.  
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1. TRYPTASE SUBFAMILY.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----

DR EMBL; M57626; AAA3998.1; -;  
DR EMBL; M57625; AAA3987.1; -;  
DR EMBL; L31853; AAA3925.1; -;  
DR EMBL; X78542; CAA5288.1; -;  
DR PIR; A38654; A38654.  
DR PIR; D35646; D35646.  
DR HSSP; P20231; IAAO.  
DR MEROPS; S01.025; -;  
DR MED; MGI:96942; Mct6.  
DR InterPro; IPR001314; Chymotrypsin.  
DR InterPro; IPR001254; Ser\_protease\_Try.  
DR Pfam; PF00089; trypsin\_1.  
DR PRINTS; PR00722; CHYMOTRYPSIN.  
DR SMART; SM00020; TRYPSIN\_DOM.  
DR PROSITE; PS50240; TRYPSIN\_DOM.  
DR PROSITE; PS00134; TRYPSIN\_HIS.  
DR PROSITE; PS00135; TRYPSIN\_SER.  
KW Hydroxylase; Serine protease; Signal; Glycoprotein; Zymogen;  
KW Alternative splicing.  
FT SIGNAL 1 21  
FT PROPEP 22 31  
FT CHAIN 32 276  
FT ACT SITE 75 75  
FT ACT SITE 122 122  
FT ACT SITE 225 225  
FT DISULFID 60 76  
FT DISULFID 156 231  
FT DISULFID 189 212  
FT DISULFID 221 249  
FT CARBOHD 133 133  
FT VARSPHC 223 230  
FT VARSPHC 231 276  
SQ SEQUENCE 276 AA; 30927 MW; 525B2C9A04A72200 CRC64;

Query Match 18.5%; Score 102; DB 1; Length 276;  
Best Local Similarity 43.8%; Pred. No. 0.00057;  
Matches 21; Conservative 9; Mismatches 18; Indels 0; Gaps 0;

QY 46 GFDIAIWMIAQMWNIQSGIRVLSLPQPSDIPPGTGVFIYGYGRDND 93  
DB 120 GADVALLLELVVNVSTHHPISLPASSTPPGTSQWVTGMDIDND 167

RESULT 3  
TRYPT\_MERUN STANDARD; PRT; 270 AA.

AC P50342;  
DT 01-OCT-1996 (Rel. 34, Created)  
DT 01-OCT-1996 (Rel. 34, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE Mast cell tryptase precursor (EC 3.4.21.59).  
OS Meriones unguiculatus (Mongolian jird) (Mongolian gerbil).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Gerbillinae;  
OC Meriones.  
OX NCBI\_TaxID=10047;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=MGS/SEA; TISSUE=Intestine;  
RX MEDLINE=95366971; PubMed=7639711;  
RA Murakumo Y., Ide H., Itoh H., Tomita M., Kobayashi T.,  
RA Matuyama H., Horii Y., Nawa Y.;  
RT "Cloning of the cDNA encoding mast cell tryptase of Mongolian gerbil,  
RT Meriones unguiculatus, and its preferential expression in the  
RT intestinal mucosa.";  
RL Biochem. J. 309:921-926(1995).  
CC -1- CATALYTIC ACTIVITY: Preferential cleavage: Arg-|-, Lys-|-, but  
CC with more restricted specificity than trypsin.  
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1. TRYPTASE SUBFAMILY.  
CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

DR EMBL; D31789; BAA06598.1; -.  
 DR HSSP; P20231; IAAO.  
 DR MEROPS; S01.143; -.  
 DR InterPro; IPR001314; Chymotrypsin.  
 DR InterPro; IPR001254; Ser\_protease\_Try.  
 DR Pfam; PF00089; trypsin.1.  
 DR PRINTS; PR00722; CHYMOTRYPSIN.  
 DR SMART; SM00020; Tryp\_Spc; 1.  
 DR PROSITE; PS50240; TRYPSIN\_DOM; 1.  
 DR PROSITE; PS00134; TRYPSIN\_HIS; 1.  
 DR PROSITE; PS00135; TRYPSIN\_SER; 1.  
 KW Hydrolyase; Serine protease; Signal; Glycoprotein.  
 FT SIGNAL 1 25 POTENTIAL.  
 FT CHAIN 26 270 MAST CELL TRYPTASE.  
 FT ACT\_SITE 69 69 CHARGE RELAY SYSTEM (BY SIMILARITY).  
 FT ACT\_SITE 116 116 CHARGE RELAY SYSTEM (BY SIMILARITY).  
 FT ACT\_SITE 219 219 CHARGE RELAY SYSTEM (BY SIMILARITY).  
 FT DISULFID 54 70 BY SIMILARITY.  
 FT DISULFID 150 225 BY SIMILARITY.  
 FT DISULFID 183 206 BY SIMILARITY.  
 FT DISULFID 215 243 BY SIMILARITY.  
 FT CARBOHYD 46 46 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 127 127 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 228 228 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 SQ SEQUENCE 270 AA; 30166 MW; 1BE102DB86943401 CRC64;

Query Match 18.3%; Score 101; DB 1; Length 270;  
 Best Local Similarity 41.2%; Pred. No. 0.00072;  
 Matches 21; Conservative 10; Mismatches 20; Indels 0; Gaps 0;

43 TLGGDIAIVMLAQMVNLOSIGRIVSLPQSPDIPPGCGVFVIGYGRDDND 93  
 111 TONGADILALEKPNVNISSHVPSLPPASSTPPSGTLCWVTGNGINDND 161

RESULT 4  
 ID TRYA\_HUMAN STANDARD; PRT; 275 AA.  
 AC P15157; Q9U011; Q9H2Y3;  
 DT 01-APR-1990 (Rel. 14, Created)  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Alpha-tryptase precursor (EC 3.4.21.59) (Trypsin 1).  
 GN TP51.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
 OC NCBI\_Taxid=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A. (ISOFORM 1).  
 RC TISSUE=Lung;  
 RA MEDLINE=90009311; PubMed=2677049;  
 RA Miller J.S., Westin E.H., Schwartz L.B.;  
 RT "Cloning and characterization of complementary DNA for human  
 RT trypsinase";  
 RL J. Clin. Invest. 84:1188-1195(1989).  
 RN [2]  
 RP REVISIONS TO 89-93 AND 108.  
 RA Schwartz L.B.;  
 RL Submitted (MAR-1990) to the EMBL/GenBank/DBJ databases.  
 RN [3]  
 RP SEQUENCE FROM N.A. (ISOFORM 1).  
 RX MEDLINE=99121069; PubMed=9920877;  
 RA Pallao M., Fejzo M.S., Shayesteh L., Blount J.L., Caughey G.H.;  
 RT "Characterization of genes encoding known and novel human mast cell

RT trypsinases on chromosome 16p13.3";  
 RL J. Biol. Chem. 274:3355-3362(1999).  
 RN [4]  
 RP SEQUENCE FROM N.A. (ISOFORMS 1 AND 2).  
 RC TISSUE=Lung;  
 RA Wang H.W., McNeill H.P., Thomas P.S., Murphy B.N., Webster M.J.,  
 RA Hettiaratchi A., King G., Heywood G.J., Huang C., Stevens R.L.,  
 RA Hunt J.E.;  
 RT "Molecular cloning and characterization of novel human trypsinase cDNAs  
 RT and splicing variants";  
 RL Submitted (NOV-1999) to the EMBL/GenBank/DBJ databases.  
 RN [5]  
 RP SEQUENCE OF 31-50, AND PITUITARY, SEQUENCE OF 31-38.  
 RC TISSUE=Lung;  
 RX MEDLINE=87109258; PubMed=3543004;  
 RA Cromlish J.A., Seidah N.G., Marcinkiewicz M., Hamelin J., Johnson D.A.,  
 RA Chretien M.;  
 RT "Human pituitary trypsinase: molecular forms, NH2-terminal sequence,  
 RT immunocytochemical localization, and specificity with prothormone and  
 RT fluorogenic substrates";  
 RL J. Biol. Chem. 262:1363-1373(1987).  
 CC -1- FUNCTION: TRYPTASE IS THE MAJOR NEUTRAL PROTEASE PRESENT IN MAST  
 CC CELLS AND IS SECRETED UPON THE COUPLED ACTIVATION-DEGRANULATION  
 CC RESPONSE OF THIS CELL TYPE.  
 CC -1- CATALYTIC ACTIVITY: Preferential cleavage: Arg-|-, Lys-|- , but  
 CC with more restricted specificity than trypsin.  
 CC -1- SUBUNIT: HOMOTETRAMER.  
 CC -1- SUBCELLULAR LOCATION: RELEASED FROM THE SECRETORY GRANULES UPON  
 CC MAST CELL ACTIVATION.  
 CC -1- ALTERNATIVE PRODUCTS: 2 isoforms; 1 (shown here) and 2; are  
 CC produced by alternative splicing.  
 CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1. TRYPTASE SUBFAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

DR EMBL; M30038; AAA86934.1; -.  
 DR EMBL; AF098328; AAD17846.1; -.  
 DR EMBL; AF206665; AAG35695.1; -.  
 DR EMBL; AF206666; AAG35696.1; -.  
 DR PIR; A45754; A45754.  
 DR HSSP; P20231; IAAO.  
 DR MEROPS; S01.015; -.  
 DR MEROPS; S01.143; -.  
 DR Gnew; HGNC:12018; TP51.  
 DR MIM; 191080; -.  
 DR InterPro; IPR001314; Chymotrypsin.  
 DR InterPro; IPR001254; Ser\_protease\_Try.  
 DR Pfam; PF00089; trypsin.1.  
 DR PRINTS; PR00722; CHYMOTRYPSIN.  
 DR SMART; SM00020; Tryp\_Spc; 1.  
 DR PROSITE; PS50240; TRYPSIN\_DOM; 1.  
 DR PROSITE; PS00134; TRYPSIN\_HIS; 1.  
 DR PROSITE; PS00135; TRYPSIN\_SER; 1.  
 KW Hydrolyase; Serine protease; Signal; Glycoprotein; Zymogen;  
 KW Polymorphism; Alternative splicing.  
 FT SIGNAL 1 18 POTENTIAL.  
 FT PROPEP 19 30 ACTIVATION PEPTIDE.  
 FT CHAIN 31 275 ALPHA-TRYPTASE.  
 FT ACT\_SITE 74 74 CHARGE RELAY SYSTEM (BY SIMILARITY).  
 FT ACT\_SITE 121 121 CHARGE RELAY SYSTEM (BY SIMILARITY).  
 FT ACT\_SITE 224 224 CHARGE RELAY SYSTEM (BY SIMILARITY).  
 FT DISULFID 59 75 BY SIMILARITY.  
 FT DISULFID 155 230 BY SIMILARITY.  
 FT DISULFID 188 211 BY SIMILARITY.  
 FT DISULFID 220 248 BY SIMILARITY.  
 FT CARBOHYD 132 132 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 233 233 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT VARSPLIC 79 87 MISSING (IN ISOFORM 2).  
 FT VARIANT 15 15 R -> P (IN ALPHA-II).  
 FT VARIANT 221 221 /FTID=VAR 012102.  
 FT VARIANT 221 221 K -> Q (IN ALPHA-II).  
 FT CONFLICT 215 216 /FTID=VAR 012103.  
 FT SEQUENCE 275 AA; 30772 MW; B9AC4BCB91CE75 CRC64;  
 Query Match 18.1%; Score 100; DB 1; Length 275;  
 Best Local Similarity 40.7%; Pred. No. 0.0094;  
 Matches 22; Conservative 10; Mismatches 20; Indels 2; Gaps 1;  
 QY 40 IAQTLSPGDIATVLAQWNLQSGIRVSLPPSPDIPPGTGVFLVGYGRDND 93  
 DB 115 IIQT--GADIALLEBEPNVISSRVHTVLPASSTFPFGMCWVGMDVDND 166  
 RESULT 5  
 MCT7 RAT STANDARD; PRT; 273 AA.  
 AC P27435; P27436; 23, Created)  
 DT 01-AUG-1992 (Rel. 23, Last sequence update)  
 DT 01-NOV-1997 (Rel. 35, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Mast cell protease 7 precursor (EC 3.4.21.59) (RMCP-7) (Tryptase, skin).  
 GN MCP7 OR MCP7.  
 OS Rattus norvegicus (Rat).  
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
 OX NCBI\_TaxID=10116;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Sprague-Dawley;  
 RX MEDLINE=97149430; PubMed=8996238;  
 RA Lutzelschab C., Pejler G., Aveskog M., Hellman L.;  
 RT "Secretory granule proteases in rat mast cells. Cloning of 10 different serine proteases and a carboxypeptidase A from various rat mast cell populations.";  
 RL J. Exp. Med. 185:13-29 (1997).  
 RN [2]  
 RP SEQUENCE OF 29-53.  
 RC STRAIN=Sprague-Dawley; TISSUE=Skin;  
 RX MEDLINE=91242400; PubMed=2036367;  
 RA Breganza V.J., Simons W.H.;  
 RT "Trypsin from rat skin: purification and properties.";  
 RL Biochemistry 30:4997-5007 (1991).  
 RN [3]  
 RP SEQUENCE OF 29-51.  
 RC TISSUE=Breast carcinoma;  
 RX MEDLINE=92231826; PubMed=1314562;  
 RA Eto I., Grubbs C.J.;  
 RT "Separation, purification and N-terminal sequence analysis of a novel leupeptin-sensitive serine endopeptidase present in chemically induced rat mammary tumour.";  
 RL Biochem. J. 283:209-216 (1992).  
 CC -1- FUNCTION: TRYPTASE IS THE MAJOR NEUTRAL PROTEASE PRESENT IN MAST CELLS, AND IS SECRETED UPON THE COUPLED ACTIVATION-DEGRADATION RESPONSE OF THIS CELL TYPE.  
 CC -1- CATALYTIC ACTIVITY: Preferential cleavage: Arg-|-, Lys-|-, but with more restricted specificity than trypsin.  
 CC -1- SUBUNIT: HOMOTETRAMER.  
 CC -1- SUBCELLULAR LOCATION: RELEASED FROM THE SECRETORY GRANULES UPON MAST CELL ACTIVATION.  
 CC -1- TISSUE SPECIFICITY: MAST CELLS.  
 CC -1- PTM: GLYCOSYLATED (PROBABLE).  
 CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1. TRYPTASE SUBFAMILY.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; U67910; AAB48263.1; -.  
 DR PIR; A23698; A23698.  
 DR PIR; S21275; S21275.  
 DR HSSP; P20231; IAAO.  
 DR MEROPS; S01.026; -.  
 DR InterPro; IPR001314; Chymotrypsin.  
 DR InterPro; IPR001254; Ser protease\_Try.  
 DR Pfam; PF00089; trypsin\_1.  
 DR PRINTS; PR00722; CHYOTRYPsin.  
 DR SMART; SM00020; TRY\_SPC; 1.  
 DR PROSITE; PS0240; TRYPSIN\_DOM; 1.  
 DR PROSITE; PS00134; TRYPSIN\_HIS; 1.  
 DR PROSITE; PS00135; TRYPSIN\_SER; 1.  
 KM Hydrolyase; Serine protease; Glycoprotein; Zymogen; Signal;  
 KW Multigene family.  
 FT SIGNAL 1 18  
 FT PROPEP 19 28  
 FT CHAIN 29 273  
 FT ACT\_SITE 72 72  
 FT ACT\_SITE 119 119  
 FT ACT\_SITE 222 222  
 FT DISULFID 57 73  
 FT DISULFID 153 228  
 FT DISULFID 186 209  
 FT DISULFID 218 246  
 FT CARBOHD 49 49  
 FT CONFLICT 42 42  
 FT CONFLICT 49 51  
 FT SEQUENCE 273 AA; 30400 MW; 65A5BD4D279FB284 CRC64;  
 Query Match 17.9%; Score 99; DB 1; Length 273;  
 Best Local Similarity 40.7%; Pred. No. 0.0012;  
 Matches 22; Conservative 10; Mismatches 20; Indels 2; Gaps 1;  
 QY 40 IAQTLSPGDIATVLAQWNLQSGIRVSLPPSPDIPPGTGVFLVGYGRDND 93  
 DB 113 IAG--DGADIALKLTNPVNITSNHTVSLPPASSTFPFGTLGMCWVGMDVDND 164  
 RESULT 6  
 TB11 HUMAN STANDARD; PRT; 275 AA.  
 AC Q15661; Q15663; Q9H2Y4;  
 DT 16-OCT-2001 (Rel. 40, Created)  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Trypsin beta-1 precursor (EC 3.4.21.59) (Trypsin 1) (Trypsin I).  
 GN TP5B1.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC MEDLINE=90251647; PubMed=2187193;  
 RA Vanderlidge P., Ballinger S.M., Tam E.K., Goldstein S.M., Craik C.S., Caughey G.H.;  
 RT "Human mast cell tryptase: multiple cDNAs and genes reveal a multigene serine protease family.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 87:3811-3815 (1990).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC MEDLINE=99121069; PubMed=9920877;  
 RA Pallao M., Fejzo M.S., Shayeesh L., Blount J.L., Caughey G.H.;  
 RT "Characterization of genes encoding known and novel human mast cell tryptases on chromosome 16p13.3.";  
 RL J. Biol. Chem. 274:3355-3362 (1999).  
 RN [3]  
 RP SEQUENCE OF 54-275 FROM N.A. (ISOFORM 2).



RA Wang H.W., McNeil H.P., Thomas P.S., Murphy B.N., Webster M.J.,  
RA Hettiaratchi A., King G., Heywood G.J., Huang C., Stevens R.L.,  
RA Hunt J.E.;  
RT "Molecular cloning and characterization of novel human trypsin cDNAs  
and splicing variants.";  
RL Submitted (NOV-1999) to the EMBL/GenBank/DBJ databases.  
CC -1- FUNCTION: TRYPTASE IS THE MAJOR NEUTRAL PROTEASE PRESENT IN MAST  
CC CELLS AND IS SECRETED UPON THE COUPLED ACTIVATION-DEGRANULATION  
CC RESPONSE OF THIS CELL TYPE.  
CC -1- CATALYTIC ACTIVITY: Preferential cleavage: Arg-|-, Lys-|-, but  
CC with more restricted specificity than trypsin.  
CC -1- SUBUNIT: HOMOTETRAMER (BY SIMILARITY).  
CC -1- SUBCELLULAR LOCATION: RELEASED FROM THE SECRETORY GRANULES UPON  
CC MAST CELL ACTIVATION.  
CC -1- ALTERNATIVE PRODUCTS: 2 isoforms; 1 (shown here) and 2; are  
CC produced by alternative splicing.  
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1. TRYPTASE SUBFAMILY.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL; M33494; AAC83172.1; -;  
DR EMBL; M33491; AAA36778.1; -;  
DR EMBL; AF099144; AAD17860.1; -;  
DR EMBL; AF206667; AAG35697.1; -;  
DR HSSP; P20231; 1AOL.  
DR MEROPS; S01.242; -;  
DR GeneW; HGNC:12019; TPSB1.  
DR MIM; 191081; -;  
DR InterPro; IPR001314; Chymotrypsin.  
DR InterPro; IPR001254; Ser\_protease\_Try.  
DR Pfam; PF00089; trypsin; 1.  
DR PRINTS; PR00722; CHYMOTRYPSIN.  
DR SMART; SMO0020; Tryp\_Spc; 1.  
DR PROSITE; PS50240; TRYPsin\_DOM; 1.  
DR PROSITE; PS00134; TRYPsin\_HIS; 1.  
DR PROSITE; PS00135; TRYPsin\_SER; 1.  
DR HydroLase; Serine protease; Signal; Glycoprotein; Zymogen;  
KW Alternative splicing.  
FT SIGNAL 1 18  
FT PROPEP 19 30  
FT CHAIN 31 275  
FT ACT\_SITE 74 74  
FT ACT\_SITE 121 121  
FT ACT\_SITE 224 224  
FT DISULFID 59 75  
FT DISULFID 155 230  
FT DISULFID 188 211  
FT DISULFID 220 248  
FT CARBOHYD 132 132  
FT CARBOHYD 233 233  
FT VARSPIC 79 87  
SQ SEQUENCE 275 AA; 30515 MW; ADC48FDC51F37112 CRC64;  
Query Match 17.9%; Score 99; DB 1; Length 275;  
Best Local Similarity 39.6%; Pred. No. 0.0012;  
Matches 19; Conservative 11; Mismatches 10; Indels 0; Gaps 0;

DR 16-OCT-2001 (Rel. 40, Last sequence update)  
DR 15-JUN-2002 (Rel. 41, Last annotation update)  
DE Putative trypsinase delta precursor (EC 3.4.21.-) (Mast cell mMCP-7-  
DE like) (HmMCP-3-like trypsinase III).  
GN TPSP1.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
OX NCBT\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX PALLADINE=99121069; PubMed=9920877;  
RA Palladio M., Fejzo M.S., Shayezen L., Blount J.L., Caughey G.H.;  
RT "Characterization of genes encoding known and novel human mast cell  
RT trypsinases on chromosome 16p13.3".  
RL J. Biol. Chem. 274:3355-3362 (1999).  
RN [2]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=21101554; PubMed=1174199;  
RA Min H.K., Kambe N., Schwartz L.B.;  
RT "Human mouse mast cell protease 7-like trypsinase genes are  
RT pseudogenes.";  
RL J. Allergy Clin. Immunol. 107:315-321 (2001).  
RN [3]  
RP SEQUENCE OF 46-235 FROM N.A.  
RA Wang H.W., McNeil H.P., Thomas P.S., Murphy B.N., Webster M.J.,  
RA Hettiaratchi A., King G., Heywood G.J., Huang C., Stevens R.L.,  
RA Hunt J.E.;  
RT "Molecular cloning and characterization of novel human trypsin cDNAs  
RT and splicing variants.";  
RL Submitted (NOV-1999) to the EMBL/GenBank/DBJ databases.  
CC -1- FUNCTION: This seems to be the product of a pseudogene, it is  
CC probably not expressed.  
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1. TRYPTASE SUBFAMILY.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL; AF098337; AAD17845.1; ALT\_INIT.  
DR EMBL; AF318074; AAK12909.1; -;  
DR EMBL; AF206664; AAG35694.1; -;  
DR HSSP; P20231; 1AOL.  
DR MEROPS; S01.054; -;  
DR GeneW; HGNC:14118; TPSP1.  
DR InterPro; IPR001314; Chymotrypsin.  
DR InterPro; IPR001254; Ser\_protease\_Try.  
DR Pfam; PF00089; trypsin; 1.  
DR PRINTS; PR00722; CHYMOTRYPSIN.  
DR SMART; SMO0020; Tryp\_Spc; 1.  
DR PROSITE; PS50240; TRYPsin\_DOM; 1.  
DR PROSITE; PS00134; TRYPsin\_HIS; 1.  
DR PROSITE; PS00135; TRYPsin\_SER; 1.  
DR HypoLase; Serine protease; Signal;  
KW Zymogen.  
FT SIGNAL 1 18  
FT PROPEP 19 30  
FT CHAIN 31 235  
FT ACT\_SITE 74 74  
FT ACT\_SITE 121 121  
FT ACT\_SITE 224 224  
FT DISULFID 59 75  
FT DISULFID 155 230  
FT DISULFID 188 211  
FT CARBOHYD 132 132  
SQ SEQUENCE 235 AA; 25816 MW; B1F86C2A8006B22 CRC64;  
Query Match 17.8%; Score 98; DB 1; Length 235;  
Best Local Similarity 40.7%; Pred. No. 0.0013;

Matches 22; Conservative 11; Mismatches 19; Indels 2; Gaps 1;

OY 40 IAGTSGFDIAVLAQWNLQSGIRVSLPQSPDIPPGTGVFIYGRDND 93  
 DB 115 IIQT--GADIALLELEBPVNISHTHTVLPASERFPFGMCWVTGMGDVNN 166

RESULT 8  
 EL3A\_HUMAN STANDARD; PRT; 270 AA.

AC P09053; Q9BRM4; (Rel. 10, Created)  
 DT 01-MAR-1989 (Rel. 10, Created)  
 DT 01-FEB-1996 (Rel. 33, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE Elastase IIA precursor (EC 3.4.21.70) (Protease E).  
 GN ELA3.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Pancreeas;  
 RX MEDLINE=88087253; PubMed=2826474;  
 RA Tani T., Ohsumi J., Mita K., Takiguchi Y.;  
 RT "Identification of a novel class of elastase isozyme, human  
 pancreatic elastase III, by cDNA and genomic gene cloning.";  
 RL J. Biol. Chem. 263:1231-1239(1988).  
 RN [2]  
 RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.  
 RC TISSUE=Pancreeas;  
 RX MEDLINE=89034017; PubMed=2460440;  
 RA Shitazu Y., Takemura K., Yoshida H., Sato Y., Iijima H.,  
 RA Shimada Y., Mikayama T., Ozawa T., Ikeda N., Ishida A., Tamai Y.,  
 RA Matsuki S., Tanaka J., Ikenaga H., Ogawa M.;  
 RT "Molecular cloning of complementary DNA encoding one of the human  
 pancreatic protease B isozymes.";  
 RL J. Biochem. 104:259-264(1988).  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Prostate;  
 RA Strausberg R.;  
 RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.  
 CC - FUNCTION: EFFICIENT PROTEASE WITH ALANINE SPECIFICITY BUT ONLY  
 CC LITTLE ELASTOLYTIC ACTIVITY.  
 CC - CATALYTIC ACTIVITY: Preferential cleavage: Ala-Xaa. Does not  
 CC hydrolyse elastin.  
 CC - SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1. ELASTASE SUBFAMILY.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
 CC -----  
 DR EMBL; M18700; AAA66350.1; -  
 DR EMBL; M18693; AAA66350.1; JOINED.  
 DR EMBL; M18694; AAA66350.1; JOINED.  
 DR EMBL; M18695; AAA66350.1; JOINED.  
 DR EMBL; M18696; AAA66350.1; JOINED.  
 DR EMBL; M18697; AAA66350.1; JOINED.  
 DR EMBL; M18698; AAA66350.1; JOINED.  
 DR EMBL; M18699; AAA66350.1; JOINED.  
 DR EMBL; M18700; AAA66350.1; JOINED.  
 DR EMBL; D00306; BAA00212.1; -  
 DR EMBL; BC005918; AAH05918.1; -  
 DR PIR; A29934; A29934.  
 DR HSSP; P05805; 1PON.  
 DR MEROPS; S01.154; -  
 DR Genew; HGNC:15944; ELA3A.  
 DR InterPro; IPR001314; Chymotrypsin.  
 DR InterPro; IPR001254; Ser\_protease\_Try.

DR Pfam; PF00089; trypsin.1.  
 DR PRINTS; PR00722; CHYMOTRYPsin.  
 DR SMART; SM00020; TRYD\_SPC; 1.  
 DR PROSITE; PS00240; TRYPSIN\_DOM; 1.  
 DR PROSITE; PS00134; TRYPSIN\_HIS; 1.  
 DR PROSITE; PS00135; TRYPSIN\_SER; 1.  
 DR KMW HydroLase; Serine protease; Zymogen; Signal; Glycoprotein.  
 FT SIGNAL 1 15  
 FT PROPEP 16 28  
 FT CHAIN 29 270  
 FT ACT\_SITE 73 73  
 FT ACT\_SITE 123 123  
 FT ACT\_SITE 217 217  
 FT DISULFID 58 74  
 FT DISULFID 117 120  
 FT DISULFID 157 223  
 FT DISULFID 188 204  
 FT DISULFID 213 244  
 FT CARBOHYD 114 114  
 FT CONFLICT 63 63  
 FT CONFLICT 106 106  
 FT CONFLICT 174 174  
 SQ SEQUENCE 270 AA; 29474 MW; 576DDB255A4A118C CRC64;

Query Match 17.3%; Score 95.5; DB 1; Length 270;  
 Best Local Similarity 29.1%; Pred. No. 0.0029;  
 Matches 25; Conservative 16; Mismatches 34; Indels 11; Gaps 1;

OY 4 LTLRNGDQGIHHQPSGVKVAAGVWPGSCMSARQRRPIQTLSGFDIAVLAQWNLQSG 63  
 DB 90 IAVKSGPQVPIPNSEELFVHPLMNRSCVAC-----GNDIALIKLSRSLQSDA 138

OY 64 IRVLSLPQSPDIPPGTGVFIYGR 89  
 DB 139 VGLASLPAGDILPKTKTCYITGMCR 164

RESULT 9  
 MCT9\_MOUSE STANDARD; PRT; 246 AA.

AC O35164; (Rel. 36, Created)  
 DT 15-JUN-1998 (Rel. 36, Last sequence update)  
 DT 15-JUN-1998 (Rel. 36, Last annotation update)  
 DE Mast cell protease 9 precursor (EC 3.4.21.-) (MMCP-9).  
 GN MCTP9.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OX NCBI\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=BAIB/CJ;  
 RX MEDLINE=96030598; PubMed=9360993;  
 RA Hunt J.E., Friend D.S., Gurish M.F., Feyfant E., Sali A., Huang C.,  
 RA Childyail N., Stechschulte S., Austen K.F., Stevens R.L.;  
 RT "Mouse mast cell protease 9, a novel member of the chromosome 14  
 RT family of serine proteases that is selectively expressed in uterine  
 RT mast cells.";  
 RL J. Biol. Chem. 272:29158-29166(1997).  
 CC - TISSUE SPECIFICITY: SELECTIVELY EXPRESSED IN UTERINE MAST CELLS.  
 CC - SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1. GRANZYME SUBFAMILY.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
 CC -----  
 DR EMBL; AF007119; AAB65245.1; -  
 DR EMBL; AF007120; AAB65246.1; -

DR HSSP; P00770; 3RP2.  
 DR MGD; MGI:1194491; Mcpt9.  
 DR InterPro; IPR001314; Chymotrypsin.  
 DR InterPro; IPR001254; Ser\_protease\_Try.  
 DR Pfam; PF00089; trypsin\_1.  
 DR PRINTS; PR00722; CHYMOTRYPSIN.  
 DR SMART; SM00202; Tryp\_Spc; 1.  
 DR PROSITE; PS50240; TRYPSIN\_DOM; 1.  
 DR PROSITE; PS00134; TRYPSIN\_HIS; 1.  
 DR PROSITE; PS00135; TRYPSIN\_SER; 1.  
 KM Hydrolyse; Serine protease; Zymogen; Signal; Multigene family.  
 FT SIGNAL 1 18  
 FT PROPEP 19 20 ACTIVATION PEPTIDE (BY SIMILARITY).  
 FT CHAIN 21 246 MAST CELL PROTEASE 9.  
 FT ACT\_SITE 65 65 CHARGE RELAY SYSTEM (BY SIMILARITY).  
 FT ACT\_SITE 109 109 CHARGE RELAY SYSTEM (BY SIMILARITY).  
 FT ACT\_SITE 202 202 CHARGE RELAY SYSTEM (BY SIMILARITY).  
 FT DISULFID 50 66 BY SIMILARITY.  
 FT DISULFID 143 208 BY SIMILARITY.  
 FT DISULFID 174 187 BY SIMILARITY.  
 SQ SEQUENCE 246 AA; 26652 MW; C3EFC5D7B51DD01 CRC64;

Query Match 16.8%; Score 93; DB 1; Length 246;  
 Best Local Similarity 42.9%; Pred. No. 0.0049;  
 Matches 18; Conservative 7; Mismatches 17; Indels 0; Gaps 0;

Qy 48 DIAIYMLQAVNLQSGIRVLSIPQSDIPPPTGTVFVYGR 89  
 Db 109 DIVLKLKKQANLTSADVVPPLPGSPDFAKPTMCMAQWGR 150

RESULT 10  
 MCTX\_MOUSE STANDARD; PRT; 246 AA.  
 AC Q00355; 01-JUL-1993 (Rel. 26, Created)  
 DT 01-JUL-1993 (Rel. 26, Last sequence update)  
 DT 15-JUL-1998 (Rel. 36, Last annotation update)  
 DE Mast cell protease-like protein precursor (EC 3.4.21.-).  
 GN MCPTL.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OC NCBI\_Taxid=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=91107701; PubMed=1988455;  
 RA Setatin W.E., Sullivan T.P., Conder G.A., Ebrahimi A., Marcham P.,  
 RA Johnson S.S., Austen K.F., Reynolds D.S.;  
 RT "Cloning of the CDNA and gene for mouse mast cell protease 4.  
 RT Demonstration of its late transcription in mast cell subclasses and  
 RT analysis of its homology to subclass-specific neutral proteases of  
 RT the mouse and rat."  
 RL J. Biol. Chem. 266:1934-1941(1991).  
 CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1. GRANZYME SUBFAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; M57401; AAA39990.1; -.  
 DR HSSP; P00770; 3RP2.  
 DR MEROPS; S01.304; -.  
 DR MGD; MGI:102792; Mcpt1.  
 DR InterPro; IPR001314; Chymotrypsin.  
 DR InterPro; IPR001254; Ser\_protease\_Try.  
 DR Pfam; PF00089; trypsin\_1.  
 DR PRINTS; PR00722; CHYMOTRYPSIN.  
 DR SMART; SM00202; Tryp\_Spc; 1.

DR PROSITE; PS50240; TRYPSIN\_DOM; 1.  
 DR PROSITE; PS00134; TRYPSIN\_HIS; 1.  
 DR PROSITE; PS00135; TRYPSIN\_SER; 1.  
 KM Hydrolyse; Serine protease; Zymogen; Signal; Multigene family.  
 FT SIGNAL 1 18  
 FT PROPEP 19 20 ACTIVATION PEPTIDE (BY SIMILARITY).  
 FT CHAIN 21 246 MAST CELL PROTEASE-LIKE PROTEIN.  
 FT ACT\_SITE 65 65 CHARGE RELAY SYSTEM (BY SIMILARITY).  
 FT ACT\_SITE 109 109 CHARGE RELAY SYSTEM (BY SIMILARITY).  
 FT ACT\_SITE 202 202 CHARGE RELAY SYSTEM (BY SIMILARITY).  
 FT DISULFID 50 66 BY SIMILARITY.  
 FT DISULFID 143 208 BY SIMILARITY.  
 FT DISULFID 174 187 BY SIMILARITY.  
 SQ SEQUENCE 246 AA; 26774 MW; 304BA25BC2B2B2AD CRC64;

Query Match 16.8%; Score 93; DB 1; Length 246;  
 Best Local Similarity 42.9%; Pred. No. 0.0049;  
 Matches 18; Conservative 7; Mismatches 17; Indels 0; Gaps 0;

Qy 48 DIAIYMLQAVNLQSGIRVLSIPQSDIPPPTGTVFVYGR 89  
 Db 109 DIVLKLKKQANLTSADVVPPLPGSPDFAKPTMCMAQWGR 150

RESULT 11  
 TRB2\_HUMAN STANDARD; PRT; 275 AA.  
 AC P20231; 09JUL17; 09JUL17; 015664; 095827;  
 DT 01-FEB-1991 (Rel. 17, Created)  
 DT 01-FEB-1991 (Rel. 17, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Trypsin beta-2 precursor (EC 3.4.21.59) (Trypsinase 2) (Trypsinase II).  
 GN TPB2 OR TPB2.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.  
 OC NCBI\_Taxid=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A. (VARIANT BETA-2).  
 RC TISSUE=lung;  
 RX MEDLINE=90369005; PubMed=2203827;  
 RA Miller J.S., Moxley G., Schwartz L.B.;  
 RT "Cloning and characterization of a second complementary DNA for human  
 RT trypsinase."  
 RL J. Clin. Invest. 86:864-870(1990).  
 RN [2]  
 RP SEQUENCE FROM N.A. (VARIANT BETA-2).  
 RX MEDLINE=93166209; PubMed=8434231;  
 RA Blom T., Hellman L.;  
 RT "Characterization of a trypsinase mRNA expressed in the human basophil  
 RT cell line KU812."  
 RL Scand. J. Immunol. 37:203-208(1993).  
 RN [3]  
 RP SEQUENCE FROM N.A. (VARIANTS BETA-2 AND BETA-3).  
 RX MEDLINE=90251647; PubMed=2187193;  
 RA Vanderslice P., Ballinger S.M., Tam E.K., Goldstein S.M., Craik C.S.,  
 RA Caughey G.H.;  
 RT "Human mast cell trypsinase: multiple cDNAs and genes reveal a  
 RT multigene serine protease family."  
 RL Proc. Natl. Acad. Sci. U.S.A. 87:3811-3815(1990).  
 RN [4]  
 RP SEQUENCE FROM N.A. (VARIANTS BETA-2 AND BETA-3).  
 RX MEDLINE=9121069; PubMed=8920877;  
 RA Pallaro M., Fejzo M.S., Shayesteh L., Blount J.L., Caughey G.H.;  
 RT "Characterization of genes encoding known and novel human mast cell  
 RT trypsinases on chromosome 16p13.3."  
 RL J. Biol. Chem. 274:3355-3362(1999).  
 RN [5]  
 RP REVISIONS.  
 RA Pallaro M., Fejzo M.S., Shayesteh L., Blount J.L., Caughey G.H.;  
 RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.  
 RN [6]  
 RP X-RAY CRYSTALLOGRAPHY (3.0 ANGSTROMS).

```

RX MEDLINE=96180625; PubMed=9521329;
RA Pereira P.J.B., Bergner A., Macedo-Ribeiro S., Huber R.,
RA Marschner G., Fritz H., Sommerhoff C.P., Bode W.;
RT "Human beta-tryptase is a ring-like tetramer with active sites facing
RL a central pore".
RL Nature 392:306-311(1998).
RN [7]
RP X-RAY CRYSTALLOGRAPHY (3.4 ANGSTROMS).
RX MEDLINE=99432168; PubMed=10500112;
RX Sommerhoff C.P., Bode W., Pereira P.J.B., Stubbs M.T.,
RA Stuerzebecher J., Piechotka G.P., Marschner G., Bergner A.;
RT "The structure of the human beta1-tryptase tetramer: fo(u)r better or
RT worse.".
RL Proc. Natl. Acad. Sci. U.S.A. 96:10984-10991(1999).
CC -!- FUNCTION: TRYPTASE IS THE MAJOR NEUTRAL PROTEASE PRESENT IN MAST
CC CELLS AND IS SECRETED UPON THE COUPLED ACTIVATION-DEGRANULATION
CC RESPONSE OF THIS CELL TYPE.
CC -!- CATALYTIC ACTIVITY: Preferential cleavage: Arg-|- , Lys-|- , but
CC with more restricted specificity than trypsin.
CC -!- SUBUNIT: HOMOTETRAMER.
CC -!- SUBCELLULAR LOCATION: RELEASED FROM THE SECRETORY GRANULES UPON
CC MAST CELL ACTIVATION.
CC -!- POLYMORPHISM: There are two alleles, beta-II and beta-III which
CC differ by 3 residues.
CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1. TRYPTASE SUBFAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M37488; AAA51843.1; -
DR EMBL; M33492; AAA36779.1; -
DR EMBL; M33493; AAA36780.1; -
DR EMBL; S55551; AAD13876.1; -
DR EMBL; AF099143; AAD17859.2; -
DR EMBL; AF099145; AAD17857.1; -
DR EMBL; AF099146; AAD17856.1; -
DR PIR; A37193; A37193.
DR PIR; B35863; B35863.
DR PDB; 1AAO; 31-JAN-94.
DR PDB; 1AOL; 23-MAR-99.
DR MEROPS; S01.027; -.
DR MEROPS; S01.242; -.
DR Genew; HGNC:14120; TPST2.
DR InterPro; IPR001314; Chymotrypsin.
DR InterPro; IPR001254; Ser_protease_Try.
DR Pfam; PF00089; trypsin; 1.
DR PRINTS; PR00722; CHYMOTRYPSIN.
DR SMART; SM00020; Tryp_Spc; 1.
DR PROSITE; PS00240; TRYPSIN_DOM; 1.
DR PROSITE; PS00134; TRYPSIN_HIS; 1.
DR PROSITE; PS00135; TRYPSIN_SER; 1.
KW Hydrolyase; Serine protease; Signal; Glycoprotein; Zymogen;
KW Polymorphism; 3D-structure.
FT SIGNAL 1 18
FT PROPEP 19 30
FT CHAIN 31 275
FT ACT_SITE 74 74
FT ACT_SITE 121 121
FT ACT_SITE 224 224
FT DISULFID 59 75
FT DISULFID 155 230
FT DISULFID 188 211
FT DISULFID 220 248
FT CARBOHD 233 233
FT VARIANT 51 53
FT CONFLICT 132 132
FT SEQUENCE 275 AA; 30529 MW; 2B27396C51F5C7A0 CRC64;

```

```

Query Match 16.8%; Score 93; DB 1; Length 275;
Best Local Similarity 37.5%; Pred. No. 0.0056;
Matches 18; Conservative 11; Mismatches 19; Indels 0; Gaps 0;

Ox 46 GPDATVWLAQMNVNQGSRIVISLPGSPDIPPGNGVIVGVGDDND 93
Db 119 GADIALLEBPVKVSSHVTTLTPASSTFPPQGMPCWVTGMGVND 166

RESULT 12
EL3B HUMAN STANDARD; PRT; 270 AA.
ID EL3B HUMAN
AC P08861; P11423;
DT 01-NOV-1988 (Rel. 09, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Elastase IIIB precursor (EC 3.4.21.70) (Protease E).
GN ELA3B.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Pancreas;
RX MEDLINE=88087253; PubMed=2826474;
RA Tani T., Ohsumi J., Mita K., Takiguchi Y.;
RT "Identification of a novel class of elastase isozyme, human
RT pancreatic elastase III, by cDNA and genomic gene cloning.".
RL J. Biol. Chem. 263:1231-1239(1988).
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=Pancreas;
RA Strausberg R.;
RL Submitted (MAR-2001) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE OF 4-270 FROM N.A.
RC TISSUE=Pancreas;
RX MEDLINE=88000545; PubMed=3477287;
RA Shen W., Fletcher T.S., Largman C.;
RT "Primary structure of human pancreatic protease E determined by
RT sequence analysis of the cloned mRNA.".
RL Biochemistry 26:3447-3452(1987).
RN [4]
RP SEQUENCE OF 31-50.
RC TISSUE=Pancreas;
RX MEDLINE=89325560; PubMed=2753124;
RA Moulard M., Kerfelec B., Mallet B., Chapus C.;
RT "Identification of a procarboxypeptidase A-truncated protease E
RT binary complex in human pancreatic juice.".
RL FEBS Lett. 250:166-170(1989).
RN [5]
RP SEQUENCE OF 94-164, AND CARBOHYDRATE-LINKAGE SITES ASN-114.
RC TISSUE=Pancreas;
RX MEDLINE=8928996; PubMed=2737288;
RA Wendorf P., Geyer R., Sziegoleit A., Linder D.;
RT "Localization and characterization of the glycosylation site of human
RT pancreatic elastase I.".
RL FEBS Lett. 249:275-278(1989).
RN [6]
RP SEQUENCE OF EFFICIENT PROTEASE WITH ALANINE SPECIFICITY BUT ONLY
RN LITTLE ELASTOLYTIC ACTIVITY.
RN -!- CATALYTIC ACTIVITY: Preferential cleavage: Ala-|-Xaa. Does not
RN hydrolyse elastin.
RN -!- SIMILARITY: Was originally (Ref.5) thought to be elastase I.
RN -----
RN This SWISS-PROT entry is copyright. It is produced through a collaboration
RN between the Swiss Institute of Bioinformatics and the EMBL outstation -
RN the European Bioinformatics Institute. There are no restrictions on its
RN use by non-profit institutions as long as its content is in no way
RN modified and this statement is not removed. Usage by and for commercial
RN entities requires a license agreement (See http://www.isb-sib.ch/announce/

```





```

QY 4 LTLRNGDQGGIHQPSGVKVAFGVM-----PSCMSARQR-----PI 40
DB 44 VSLRRDQYV-RHGGGSLIHQWVLTAAHCIPGLQSPDRVQLRQHLYVYDRLPI 102
QY 41 AQTLE-----SGFDIAVLAQMVLQSGIRVLSLPQSPDIPPGTGVEIVGRDD 91
DB 103 SRVIFPHRYVWVNGADIALQLLEPVSISRHVQPVTLPPASSTPPESQCVGWGDVD 162
QY 92 NDR 94
DB 163 NGR 165

RESULT 16
TRYPT_PIG STANDARD; PRT; 275 AA.
AC Q9N2DI;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Trypsinase precursor (EC 3.4.21.59).
GN MCTV.
OS Sus scrofa (pig).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
OX NCBI_TaxID=9823;
[1]
RP SEQUENCE FROM N.A.
RC TISSUE=Lung;
RX MEDLINE=20285343; PubMed=10824103;
RA Chen Y., Shioita M., Ohuchi M., Towatari T., Tashiro J., Murakami M.,
RA Yano M., Yang B., Kido H.;
RT "Mast cell tryptase from pig lungs triggers infection by pneumotropic
RT Sendai and influenza A viruses. Purification and characterization.";
RL Eur. J. Biochem. 267:3189-3197(2000).
CC -1- FUNCTION: TRYPTASE IS THE MAJOR NEUTRAL PROTEASE PRESENT IN MAST
CC CELLS AND IS SECRETED UPON THE COUPLED ACTIVATION-DEGRANULATION
CC RESPONSE OF THIS CELL TYPE.
CC -1- CATALYTIC ACTIVITY: Preferential cleavage: Arg-|- , Lys-|- , but
CC with more restricted specificity than trypsin.
CC -1- SUBUNIT: HOMOTETRAMER (By similarity).
CC -1- SUBCELLULAR LOCATION: RELEASED FROM THE SECRETORY GRANULES UPON
CC MAST CELL ACTIVATION.
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1. TRYPTASE SUBFAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AB036552; BAA93614.1; -
CC HSSP; P20231; IAAO.
CC MEROPS; S01.143; -.
DR InterPro; IPR001314; Chymotrypsin.
DR InterPro; IPR001254; Ser_protease_Try.
DR Pfam; PF00089; trypsin.1.
DR PRINTS; PR00722; CHYMOTRYPSIN.
DR SMART; SM00020; Tryp_Spc; 1.
DR PROSITE; PS50240; TRYPSIN_DOM; 1.
DR PROSITE; PS00134; TRYPSIN_HIS; 1.
DR PROSITE; PS00135; TRYPSIN_SER; 1.
KW Hydroxylase; Serine protease; Signal; Glycoprotein; Zymogen.
FT SIGNAL 1 20
FT PROPEP 21 30 ACTIVATION PEPTIDE (BY SIMILARITY).
FT CHAIN 31 275 TRYPTASE.
FT ACT_SITE 74 74 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE 121 121 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE 224 224 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT DISULFID 59 75 BY SIMILARITY.
FT DISULFID 155 230 BY SIMILARITY.

```

```

FT DISULFID 188 211 BY SIMILARITY.
FT DISULFID 220 248 BY SIMILARITY.
FT CARBOHYD 132 132 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 233 233 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 275 AA; 30439 MW; ACC582647FCCB973 CRC64;

Query Match 16.0%; Score 88.5; DB 1; Length 275;
Best Local Similarity 26.4%; Pred. No. 0.017;
Matches 24; Conservative 21; Mismatches 27; Indels 19; Gaps 3;

QY 3 FLTLRNGDQGGIHQPSGVKVAFGVM-----PSCMSARQRPIACTLSGFDIAVLAQM 57
DB 85 YIRVQLGEQH-LYYODRLVLSRIIVHNYV-----DEVNGADIALLELDP 130
QY 58 VNLQSGIRVLSLPQSPDIPPGTGVEIVGYG 88
DB 131 VNLSSHVPVTLPPASSTPPKGTGRCVWGWG 161

RESULT 17
E1L_BOVIN STANDARD; PRT; 266 AA.
AC Q281S3;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Elastase 1 precursor (EC 3.4.21.36).
GN ELA1.
OS Bos taurus (Bovine).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
OC Bovidae; Bovinae; Bos.
OX NCBI_TaxID=9913;
[1]
RP SEQUENCE FROM N.A.
RC STRAIN=Holstein-Friesian; TISSUE=Pancreas;
RX MEDLINE=98079203; PubMed=9418008;
RA Gestin M., le Huereu-Luron I., Wicker-Planquart C., le Dren G.,
RA Chaix J.C., Puigserver A., Guilloteau P., et al;
RT "Bovine pancreatic preproelastases I and II: comparison of nucleotide
RT and amino acid sequences and tissue specific expression.";
RL Comp. Biochem. Physiol. 118B:181-187(1997).
CC -1- FUNCTION: ACTS UPON ELASTIN.
CC -1- CATALYTIC ACTIVITY: Hydrolysis of proteins, including elastin.
CC Preferential cleavage: Ala-|-Xaa.
CC -1- SUBCELLULAR LOCATION: Secreted.
CC -1- TISSUE SPECIFICITY: PANCREAS.
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; M80838; AAA98525.1; -
CC HSSP; P00772; IELG.
CC MEROPS; S01.153; -.
DR InterPro; IPR001314; Chymotrypsin.
DR InterPro; IPR001254; Ser_protease_Try.
DR Pfam; PF00089; trypsin.1.
DR PRINTS; PR00722; CHYMOTRYPSIN.
DR SMART; SM00020; Tryp_Spc; 1.
DR PROSITE; PS50240; TRYPSIN_DOM; 1.
DR PROSITE; PS00134; TRYPSIN_HIS; 1.
DR PROSITE; PS00135; TRYPSIN_SER; 1.
KW Hydroxylase; Serine protease; Zymogen; Signal.
FT SIGNAL 1 16
FT PROPEP 17 26 ACTIVATION PEPTIDE (BY SIMILARITY).
FT CHAIN 27 266 ELASTASE 1.
FT ACT_SITE 71 71 CHARGE RELAY SYSTEM (BY SIMILARITY).

```





```
DR EMBL; L00116; AAA98811.1; JOINED.
DR PIR; A00960; ELRT1.
DR HSSP; P00772; IELG.
DR MEROPS; S01.153; -.
DR InterPro; IPR001314; Chymotrypsin.
DR InterPro; IPR001254; Ser_protease_Try.
DR Pfam; PF00089; trypsin; 1.
DR PRINTS; PR00722; CHYMOTRYPSIN.
DR SMART; SM00020; Tryp_Spc; 1.
DR PROSITE; PS50240; TRYPSIN_DOM; 1.
DR PROSITE; PS00134; TRYPSIN_HIS; 1.
DR PROSITE; PS00135; TRYPSIN_SER; 1.
DR HydroLase; Serine protease; Zymogen; Signal.
KM SIGNAL
FT PROPEP 1 16
FT CHAIN 17 26 ACTIVATION PEPTIDE.
FT ACT_SITE 27 266 ELASTASE 1.
FT ACT_SITE 71 71 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE 119 119 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE 214 214 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT DISULFID 56 72 BY SIMILARITY.
FT DISULFID 153 220 BY SIMILARITY.
FT DISULFID 184 200 BY SIMILARITY.
FT DISULFID 210 240 BY SIMILARITY.
FT CONFLICT 104 104 M -> V (IN REF. 2).
FT CONFLICT 108 108 T -> N (IN REF. 2).
FT CONFLICT 244 244 K -> R (IN REF. 2).
FT CONFLICT 266 266 T -> N (IN REF. 2).
SQ SEQUENCE 266 AA; 28976 MW; 5A56FE8FCF1AEDA CRC64;

Query Match 15.4%; Score 85; DB 1; Length 266;
Best Local Similarity 27.6%; Pred. No. 0.041;
Matches 27; Conservative 19; Mismatches 32; Indels 20; Gaps 3;

QY 4 LTRNGDQGIHHSQGVK--VAPGYPSCMSARQR---PIAQ-TLSEGDIAIVMLAQMVN 63
DB 88 LSQNDGTQYVSQK--IMVHPTMNSNNVA-----GYDIALRLAQSVTLNLY 134
QY 64 IRVISTLPQSDIPPGTGVFIYVGRDNDPRSRKNG 101
DB 135 VQLAVLPQEGTILANNPCYINGWGR-----TRING 165

RESULT 20
MPN_HUMAN STANDARD; PRT; 290 AA.
AC QEQQR3;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Marapsin precursor (EC 3.4.21.-).
GN MPN.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Fortunato M., Dando P.M., Rawlings N.D., Barrett A.J.;
RT "Cloning, sequencing and expression of marapsin, a human serine
RT proteinase."
RL Submitted (MAR-2001) to the EMBL/Genbank/DBJ databases.
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AJ306593; CAC35467.1; -.
DR HSSP; P00734; IUVS.
```

```
DR MEROPS; S01.074; -.
DR InterPro; IPR001314; Chymotrypsin.
DR InterPro; IPR001254; Ser_protease_Try.
DR Pfam; PF00089; trypsin; 1.
DR PRINTS; PR00722; CHYMOTRYPSIN.
DR SMART; SM00020; Tryp_Spc; 1.
DR PROSITE; PS50240; TRYPSIN_DOM; 1.
DR PROSITE; PS00134; TRYPSIN_HIS; 1.
DR PROSITE; PS00135; TRYPSIN_SER; 1.
DR HydroLase; Serine protease; Zymogen; Signal; Glycoprotein.
FT SIGNAL 1 22
FT PROPEP 23 34 POTENTIAL.
FT CHAIN 35 290 MARAP SIN.
FT ACT_SITE 35 277 SERINE PROTEASE.
FT ACT_SITE 75 75 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE 124 124 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE 229 229 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT DISULFID 60 76 BY SIMILARITY.
FT DISULFID 158 235 BY SIMILARITY.
FT DISULFID 191 214 BY SIMILARITY.
FT DISULFID 225 253 BY SIMILARITY.
FT CARBOHYD 55 55 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 79 79 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 290 AA; 31940 MW; 67BDC93EC70BFF7B CRC64;

Query Match 15.1%; Score 83.5; DB 1; Length 290;
Best Local Similarity 32.6%; Pred. No. 0.065;
Matches 29; Conservative 11; Mismatches 38; Indels 9; Gaps 4;

QY 7 RNQDQGIHHSQGVK--VAPGYPSCMSARQR---PIAQ-TLSEGDIAIVMLAQMVN 59
DB 78 RNTSETSLYQLILGARQLVQPG--PHMYARVQVESNPLYQGTASADVALVELAEVVP 135
QY 60 LOSGRVISTLPQSDIPPGTGVFIYVGR 88
DB 136 FTVYILPVLCPDPVSIVFETGMNCWVTGNG 164

RESULT 21
MCT7_MOUSE STANDARD; PRT; 273 AA.
ID MCT7_MOUSE
AC Q02844;
DT 01-JUL-1993 (Rel. 26, Created)
DT 01-JUL-1993 (Rel. 26, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Mast cell protease 7 precursor (EC 3.4.21.59) (MMCP-7) (TRYPTASE).
GN MCP7.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A. (TRUNCATED ISOFORM).
RA STRAIN=DBA/2;
RX MEDLINE=93087489; Pubmed=1454796;
RA McNeil H.P., Reynolds D.S., Schiller V., Ghildyal N., Gurtley D.S.,
RA Austen K.F., Stevens R.L.;
RT "Isolation, characterization, and transcription of the gene encoding
RT mouse mast cell protease 7."
RL Proc. Natl. Acad. Sci. U.S.A. 89:11174-11178(1992).
RN [2]
RP SEQUENCE FROM N.A. (TRUNCATED ISOFORM).
RA STRAIN=C57BL/6J;
RX MEDLINE=96162035; Pubmed=8576265;
RA Hunt J.E., Stevens R.L., Austen K.F., Zhang J., Xia Z.,
RA Ghildyal N.;
RT "Natural disruption of the mouse mast cell protease 7 gene in the
RT C57BL/6 mouse."
RL J. Biol. Chem. 271:2851-2855(1996).
CC -1- FUNCTION: TRYPTASE IS THE MAJOR NEUTRAL PROTEASE PRESENT IN MAST
CC CELLS. AND IS SECRETED UPON THE COUPLED ACTIVATION-DEGRADATION
CC RESPONSE OF THIS CELL TYPE.
CC -1- CATALYTIC ACTIVITY: Preferential cleavage: Arg-|-, Lys-|-, but
```

CC with more restricted specificity than trypsin.  
CC -1- ALTERNATIVE PRODUCTS: 2 isoforms; a long form (shown here) and a  
CC truncated form; are produced by alternative splicing. The  
CC alternative splicing event is due to a G to A point mutation at  
CC the exon 2/intron 2 splice site and causes loss of protein  
CC expression. The alternatively spliced transcript is only found in  
CC C57BL/6 mouse.  
CC -1- DEVELOPMENTAL STAGE: IS NOT EXPRESSED IN MATURE SEROSAL OR  
CC MUCOSAL MAST CELLS AND IS EXPRESSED ONLY TRANSIENTLY AT AN EARLY  
CC STAGE OF IN VITRO MAST CELL DIFFERENTIATION.  
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1. TRYPTASE SUBFAMILY.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL; L00654; AAA39993.1; -;  
DR EMBL; L00653; AAA39992.1; -;  
DR EMBL; U42405; AAA97874.1; -;  
DR EMBL; U42406; AAA97875.1; -;  
DR PIR; A47246; A47246.  
DR HSSP; P20231; IAAO.  
DR MEROPS; S01.026; -;  
DR MGD; MGI:96943; Mcp7.  
DR InterPro; IPR001314; Chymotrypsin.  
DR InterPro; IPR001254; Ser\_protease\_Try.  
DR Pfam; PF00089; trypsin; 1.  
DR PRINTS; PR00722; CHYMOTRYPSIN.  
DR SMART; SM00020; Tryp\_Spc; 1.  
DR PROSITE; PS00240; TRYPSIN\_DOM; 1.  
DR PROSITE; PS00134; TRYPSIN\_HIS; 1.  
DR PROSITE; PS00135; TRYPSIN\_SER; 1.  
KM Hydrolase; Serine protease; Signal; Zymogen; Alternative splicing;  
KW Glycoprotein.  
FT SIGNAL 1 18 POTENTIAL.  
FT PROPEP 19 28 ACTIVATION PEPTIDE (POTENTIAL).  
FT CHAIN 29 273 MAST CELL PROTEASE 7.  
FT ACT\_SITE 72 72 CHARGE RELAY SYSTEM (BY SIMILARITY).  
FT ACT\_SITE 119 119 CHARGE RELAY SYSTEM (BY SIMILARITY).  
FT ACT\_SITE 222 222 CHARGE RELAY SYSTEM (BY SIMILARITY).  
FT DISULFID 57 73 BY SIMILARITY.  
FT DISULFID 153 228 BY SIMILARITY.  
FT DISULFID 186 209 BY SIMILARITY.  
FT DISULFID 218 246 BY SIMILARITY.  
FT CARBOHYD 49 49 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 130 130 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT VARSPLIC 44 46 VSL -> GCC (IN TRUNCATED ISOFORM).  
FT VARSPLIC 47 273 MISSING (IN TRUNCATED ISOFORM).  
SQ SEQUENCE 273 AA; 30337 MW; 50BCB4765294205E CRC64;  
Query Match 15.0%; Score 83; DB 1; Length 273;  
Best Local Similarity 36.2%; Pred. No. 0.069;  
Matches 17; Conservative 9; Mismatches 21; Indels 0; Gaps 0;  
OY 46 GPDIVIMLAQMVNLSGSRVISTLPSPDIPPGTCGVFTVIGGRDN 92  
DB 117 GADIALMLKLTNPNTISDYHVPVLPASSETFPGTLCVWTGMGNIDN 163  
RESULT 22  
ST14\_HUMAN STRAND; PRT; 855 AA.  
AC Q9YSY6; Q9H3S0; Q9HCA3; Q9BS01; Q9HB36;  
DT 16-OCT-2001 (Rel. 40, Created)  
DT 16-OCT-2001 (Rel. 40, Last sequence update)  
DT 15-JUN-2002 (Rel. 41, Last annotation update)  
DE Suppressor of tumorigenicity 14 (SC3.4.21.-) (Matrilinease) (Membrane-  
DE type serine protease 1) (MT-SPI) (Proctamin) (Serine protease TADG-15)  
DE (Tumor associated differentially-expressed gene-15 protein).

GN ST14 OR PRSS14 OR SNCL9 OR TADG15.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=99303581; PubMed=10373424;  
RA Lin C.Y., Anders J., Johnson M., Sang Q.A., Dickson R.B.;  
RT "Molecular cloning of cDNA for matrilinease, a matrix-degrading serine  
RT protease with trypsin-like activity";  
RL J. Biol. Chem. 274:18231-18236(1999).  
RN [2]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=99432178; PubMed=10500122;  
RA Takeuchi T., Shuman M.A., Craik C.S.;  
RT "Reverse biochemistry: Use of macromolecular protease inhibitors to  
RT dissect complex biological processes and identify a membrane-type  
RT serine protease in epithelial cancer and normal tissue";  
RL Proc. Natl. Acad. Sci. U.S.A. 96:11054-11061(1999).  
RN [3]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Prostate;  
RA Yamaguchi N., Mitsui S.;  
RT "Molecular cloning of a novel transmembrane serine protease expressed  
RT in human prostate";  
RL Submitted (JUL-1999) to the EMBL/GenBank/DBJ databases.  
RN [4]  
RP SEQUENCE FROM N.A.  
RA Tanimoto H., Underwood L.J., Wang Y., Shigemasa K., Parmley T.H.,  
RA O'Brien T.J.;  
RL Submitted (APR-1998) to the EMBL/GenBank/DBJ databases.  
RN [5]  
RP SEQUENCE OF 327-855 FROM N.A.  
RC TISSUE=Muscle;  
RA Strausberg R.;  
RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.  
RN [6]  
RP SEQUENCE OF 340-664 FROM N.A.  
RA Cao J., Fan W., Zheng S.;  
RL "Genomic analysis of a novel human serine protease SNCL9";  
RL Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.  
RN [7]  
RP CHARACTERIZATION.  
RC TISSUE=Milk;  
RX MEDLINE=99303582; PubMed=10373425;  
RA Lin C.Y., Anders J., Johnson M., Dickson R.B.;  
RT "Purification and characterization of a complex containing matrilinease  
RT and a Kunitz-type serine protease inhibitor from human milk";  
RL J. Biol. Chem. 274:18237-18242(1999).  
CC -1- FUNCTION: DEGRADATES EXTRACELLULAR MATRIX. PROPOSED TO PLAY A ROLE  
CC IN BREAST CANCER INVASION AND METASTASIS. EXHIBITS TRYPSIN-LIKE  
CC ACTIVITY AS DEFINED BY CLEAVAGE OF SYNTHETIC SUBSTRATES WITH ARG  
CC OR LYS AS THE P1 SITE.  
CC -1- SUBCELLULAR LOCATION: Type II membrane protein (Probable).  
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1.  
CC -1- SIMILARITY: CONTAINS 4 LDL-RECEPTOR CLASS A DOMAINS.  
CC -1- SIMILARITY: CONTAINS 2 CUB DOMAINS.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL; AF118224; AAD42765.2; -;  
DR EMBL; AF113086; AAF00109.1; -;  
DR EMBL; AB030036; BAB020376.1; -;  
DR EMBL; AF057145; AAG15395.1; -;  
DR EMBL; BC005826; AAG05826.1; -;  
DR EMBL; AF283256; AAG13949.1; -;

```

DR HSSE; P00763; IDPO.
DR Genew; HGNC:11344; ST14.
DR MIM; 606797; -.
DR MEROPS; S01.302; -.
DR InterPro; IPR000859; CUB_domain.
DR InterPro; IPR001314; Chymotrypsin.
DR InterPro; IPR00172; LDL_recept_A.
DR InterPro; IPR001254; Ser_protease_Try.
DR Pfam; PF00057; ldl_recept_a; 4.
DR Pfam; PF00089; trypsin; 1.
DR Pfam; PF00431; CUB_2.
DR PRINTS; PR00722; CHYMOTRYPSIN.
DR PRINTS; PR00261; LDLRECEPTOR.
DR SMART; SM00042; CUB; 2.
DR SMART; SM00192; LDLA; 3.
DR SMART; SM00020; Tryp_spec; 1.
DR PROSITE; PS01180; CUB; 2.
DR PROSITE; PS01209; LDLRA_1; 2.
DR PROSITE; PS50068; LDLRA_2; 4.
DR PROSITE; PS50240; TRYPsin DOM; 1.
DR PROSITE; PS00134; TRYPsin HIS; 1.
DR PROSITE; PS00135; TRYPsin SER; 1.
KW Signal-anchor; Glycoprotein; Hydrolase; Serine protease;
KM Transmembrane; Repeat.
FT DOMAIN 1 55
FT TRANSMEM 56 76
FT FT
FT FT
FT DOMAIN 77 855
FT DOMAIN 214 334
FT DOMAIN 340 447
FT DOMAIN 452 487
FT DOMAIN 487 524
FT DOMAIN 524 560
FT DOMAIN 566 603
FT DOMAIN 615 854
FT ACT_SITE 656 656
FT ACT_SITE 711 711
FT ACT_SITE 805 805
FT CARBOHYD 109 109
FT CARBOHYD 302 302
FT CARBOHYD 485 485
FT CARBOHYD 772 772
FT CONFLICT 327 329
FT CONFLICT 381 381
FT CONFLICT 674 674
SQ SEQUENCE 855 AA; 94769 MW; 26143132C01F99C9 CRC64;

Query Match 14.9%; Score 82.5; DB 1; Length 855;
Best Local Similarity 27.1%; Pred. No. 0.29;
Matches 23; Conservative 15; Mismatches 28; Indels 19; Gaps 3;

QY 13 GIHHQPSGVKVAQPMSCMSARQRPPIAQTSS-----GPDIAVYMLAQMNVLQSG 63
DB 677 GIHHDSQ--RSAPGV-----QERRLRKRISHPFNDFPDYDIALLEKPAEYSSM 726
QY 64 IRVISLPQSDIPPGTGVFIYGYG 88
DB 727 VRPICLPDASHVFPAGKAIWVGWG 751

RESULT 23
MCT2_RAT
ID_MCT2_RAT STANDARD; PRT; 247 AA.
AC P00770;
DT 21-JUL-1986 (Rel. 01, Created)
DT 01-JAN-1988 (Rel. 06, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Mast cell protease II precursor (EC 3.4.21.-) (RMCP-II) (RMCP-2)
DE (Group-specific protease).
OS Rattus norvegicus (rat).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;

```

```

RN [1]
RP SEQUENCE FROM N.A.
RP MEDLINE=87165980; PubMed=3549719;
RA Bentley P.N., Yin F.H., Leder P.;
RT "Cloning of the mast cell protease, RMCP II. Evidence for
RT cell-specific expression and a multi-gene family.";
RL J. Biol. Chem. 262:5377-5384(1987).
RN [2]
RP SEQUENCE OF 21-244.
RP MEDLINE=78124137; PubMed=629933;
RA Woodbury R.G., Katunuma N., Kobayashi K., Titani K., Neurath H.;
RT "Covalent structure of a group-specific protease from rat small
RT intestine. Appendix: crystallographic data for a group specific
RT protease from rat intestine.";
RL Biochemistry 17:811-819(1978).
RN [3]
RP X-RAY CRYSTALLOGRAPHY (1.9 ANGSTROMS).
RP MEDLINE=89166471; PubMed=323198;
RA Remington S.J., Woodbury R.G., Reynolds R.A., Matthews B.W.,
RA Neurath H.;
RT "The structure of rat mast cell protease II at 1.9-A resolution.";
RL Biochemistry 27:8097-8105(1988).
CC -1- FUNCTION: THIS ENZYME, ISOLATED FROM SMALL INTESTINE, SPECIFICALLY
CC INACTIVATES THE APO FORMS OF A CERTAIN GROUP OF INTRACELLULAR
CC PYRIDOXAL PHOSPHATE-REQUIRING ENZYMES. IT HAS CHYMOTRYPSIN-LIKE
CC SPECIFICITY TOWARDS SMALL SUBSTRATES.
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1. GRANYME SUBFAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; J02712; AAA66284.1; -.
DR PIR; A00957; PRRYG.
DR PIR; A29548; A29548.
DR PDB; 3RP2; 09-JUN-89.
DR MEROPS; S01.141; -.
DR InterPro; IPR001314; Chymotrypsin.
DR InterPro; IPR001254; Ser_protease_Try.
DR Pfam; PF00089; trypsin; 1.
DR PRINTS; PR00722; CHYMOTRYPSIN.
DR SMART; SM00020; TRYPsin; 1.
DR PROSITE; PS50240; TRYPsin DOM; 1.
DR PROSITE; PS00134; TRYPsin HIS; 1.
DR PROSITE; PS00135; TRYPsin SER; 1.
KW Hydrolase; Serine protease; 3d-structure; Zymogen; Signal;
KW Multigene family.
FT SIGNAL 1 18
FT PROPEP 19 20
FT CHAIN 21 247
FT ACT_SITE 65 65
FT ACT_SITE 109 109
FT ACT_SITE 202 202
FT DISULFID 50 66
FT DISULFID 143 208
FT DISULFID 174 187
FT CONFLICT 238 238
FT STRAND 22 22
FT STRAND 22 22
FT STRAND 25 26
FT TURN 29 30
FT TURN 33 34
FT STRAND 35 41
FT TURN 43 44
FT STRAND 47 54
FT STRAND 59 62
FT HELIX 64 66
FT STRAND 70 75
FT STRAND 79 79
FT TURN 80 81

```

```
FT TURN 85 86
FT STRAND 88 97
FT TURN 99 100
FT STRAND 111 115
FT STRAND 122 122
FT TURN 123 124
FT STRAND 125 125
FT TURN 133 134
FT TURN 139 140
FT STRAND 142 147
FT STRAND 150 152
FT TURN 153 154
FT STRAND 155 157
FT STRAND 160 160
FT STRAND 162 169
FT HELIX 171 173
FT TURN 174 178
FT TURN 182 184
FT STRAND 185 188
FT TURN 191 192
FT STRAND 196 196
FT TURN 199 203
FT STRAND 205 208
FT TURN 209 210
FT STRAND 211 218
FT TURN 221 222
FT STRAND 227 231
FT HELIX 232 243
SQ SEQUENCE 247 AA; 27101 MW; 051988042a97a47e CRC64;
```

Query Match 14.8%; Score 81.5; DB 1; Length 247;  
Best Local Similarity 36.0%; Pred. No. 0.09;  
Matches 18; Conservative 9; Mismatches 22; Indels 1; Gaps 1;

```
Oy 48 DIAIVMLAQMVNLQSGIRVLSLPQSDIPRPGTGVFIVGYGRDDNDRDS 97
Db 109 DIMLTKELKVELTPAVNVPLPSPDFTHPAMCMAGMGK-TGVDRPT 157
```

RESULT 24  
MCT4 RAT  
ID MCT4 RAT STANDARD; PRT; 246 AA.

```
AC P97592;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Mast cell protease IV precursor (EC 3.4.21.-) (RMCP-IV) (RMCP-4).
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxId=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Mistar;
RX MEDLINE=97149430; PubMed=8996238;
RA Lutzelschab C., Pejler G., Aveskog M., Hellman L.;
RT "Secretory granule proteases in rat mast cells. Cloning of 10
RT different serine proteases and a carboxypeptidase A from various rat
RT mast cell populations." (1997).
RL J. Exp. Med. 185:13-29(1997).
CC -1- SUBCELLULAR LOCATION: Secretory granules.
CC -1- TISSUE SPECIFICITY: MAST CELLS.
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1. GRANZYME SUBFAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; U67907; AAB48260.1; -;
```

```
DR HSSP; P00770; 3RP2.
DR MEROPS; S01.005; -.
DR InterPro; IPR001314; Chymotrypsin.
DR InterPro; IPR001254; Ser Protease_Try.
DR Pfam; PF00089; trypsin; 1.
DR PRINTS; PR00722; CHYMOTRYPSIN.
DR SMART; SM00020; TRYP_SPE; 1.
DR PROSITE; PS50240; TRYPSIN_DOM; 1.
DR PROSITE; PS00134; TRYPSIN_HIS; 1.
DR PROSITE; PS00135; TRYPSIN_SER; FALSE NEG.
KW Hydrolyase; Serine protease; Signal; Zymogen; Multigene family.
FT SIGNAL 1 18
FT PROPEP 19 20
FT CHAIN 21 246
FT ACT_SITE 65 65
FT ACT_SITE 109 109
FT ACT_SITE 202 202
FT DISULFID 50 66
FT DISULFID 143 208
FT DISULFID 174 187
SQ SEQUENCE 246 AA; 27042 MW; 412FF51D05098778 CRC64;
```

Query Match 14.7%; Score 81; DB 1; Length 246;  
Best Local Similarity 40.4%; Pred. No. 0.1;  
Matches 23; Conservative 7; Mismatches 25; Indels 2; Gaps 2;

```
Oy 35 RQRRIPIQTL-SGF-DIAIVMLAQMVNLQSGIRVLSLPQSDIPRPGTGVFIVGYGR 89
Db 94 KQIFPLKXNLNLSNFRDMLLTKELQKAVLTPSVNVPLPQSSDIIRKPGTMCLAGMGQ 150
```

RESULT 25  
CORI MOUSE  
ID CORI MOUSE STANDARD; PRT; 1113 AA.

```
AC Q92319;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Atrial natriuretic peptide-converting enzyme (EC 3.4.21.-) (pro-ANP-
DE converting enzyme) (Corin) (Low density lipoprotein receptor related
DE protein 4).
GN CRN OR LRp4.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxId=10090;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=98429596; PubMed=9756624;
RA Tomita Y., Kim D.-H., Magoori K., Fujino T., Yamamoto T.T.;
RT "A novel low-density lipoprotein receptor-related protein with type II
RT membrane protein-like structure is abundant in heart.";
RL J. Biochem. 124:784-789(1998).
CC -1- FUNCTION: CONVERTS PRO-ANP TO ANP. CLEAVES PRO-ANP SPECIFICALLY
CC BETWEEN ARG-122 AND SER-123 (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: Type II membrane protein.
CC -1- TISSUE SPECIFICITY: HIGHLY EXPRESSED IN HEART.
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1.
CC -1- SIMILARITY: CONTAINS 7 LBD-RECEPTOR CLASS A DOMAINS.
CC -1- SIMILARITY: CONTAINS 2 FRIZZLED (FZ) DOMAINS.
CC -1- SIMILARITY: CONTAINS 1 SRCR DOMAIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AB013874; BAA34371.1; -.
DR HSSP; P00763; IDPO.
DR MEROPS; S01.019; -;
```

```

DR MGD, MGI:1349451, Lrp4.
DR InterPro; IPR001314; Chymotrypsin.
DR InterPro; IPR000024; Fz domain.
DR InterPro; IPR002172; LDL_recept_A.
DR InterPro; IPR001254; Ser_protease_Try.
DR InterPro; IPR001190; Strc_receptor.
DR Pfam; PF00057; LDL_recept_a; 6.
DR Pfam; PF00089; trypsin; 1.
DR PRINTS; PF01392; Fz; 2.
DR PRINTS; PR00722; CHYMOTRYPSIN.
DR PRINTS; PR00261; LDLRECEPTOR.
DR SMART; SM00063; FRI; 2.
DR SMART; SM00192; LDLa; 7.
DR SMART; SM00202; SR; 1.
DR SMART; SM00020; Tryp_Spc; 1.
DR PROSITE; PS50038; Fz; 2.
DR PROSITE; PS01209; LDLRA_1; 6.
DR PROSITE; PS50068; LDLRA_2; 7.
DR PROSITE; PS00420; SRCR_1; FALSE_NEG.
DR PROSITE; PS50287; SRCR_2; 1.
DR PROSITE; PS50240; TRYPSIN_DOM; 1.
DR PROSITE; PS00134; TRYPSIN_HIS; FALSE_NEG.
DR PROSITE; PS00135; TRYPSIN_SER; 1.
DR HydroLase; Serine protease; Transmembrane; Signal-anchor;
KW Glycoprotein; Repeat.
FT DOMAIN 1 112 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 1 133 SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN)
FT 1 112 (POTENTIAL).
FT 1 133 EXTRACELLULAR (POTENTIAL).
FT DOMAIN 134 113 Fz 1.
FT 201 327 LDL-RECEPTOR CLASS A 1.
FT DOMAIN 336 372 LDL-RECEPTOR CLASS A 2.
FT 373 408 LDL-RECEPTOR CLASS A 2.
FT 409 445 LDL-RECEPTOR CLASS A 3.
FT DOMAIN 446 483 LDL-RECEPTOR CLASS A 4.
FT 518 641 Fz 2.
FT 647 682 LDL-RECEPTOR CLASS A 5.
FT DOMAIN 683 721 LDL-RECEPTOR CLASS A 6.
FT 722 757 LDL-RECEPTOR CLASS A 7.
FT DOMAIN 758 853 SRCR.
FT 869 1113 SERINE PROTEASE.
FT ACT_SITE 910 910 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE 959 959 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE 1052 1052 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT DISULFID 857 979 BY SIMILARITY.
FT DISULFID 895 911 BY SIMILARITY.
FT DISULFID 1022 1037 BY SIMILARITY.
FT DISULFID 1048 1077 BY SIMILARITY.
FT CARBOHYD 147 147 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 202 202 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 208 208 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 298 298 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 317 317 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 373 373 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 411 411 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 444 444 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 481 481 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 519 519 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 537 537 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 635 635 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 719 719 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 765 765 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 828 828 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 970 970 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1089 1089 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 1113 AA; 122984 MW; B845B2C5F20DB8EC CRC64;
Query Match 14.7%; Score 81; DB 1; Length 1113;
Best Local Similarity 28.0%; Pred. No. 0.57;
Matches 23; Conservative 14; Mismatches 39; Indels 6; Gaps 2;

```

```

Cy 71 QPSDIPPGTGIVGYGRDN 92
Db 982 SPEEYLEDYCYINGWGMGN 1003

RESULT 26
HFFL_HUMAN STANDARD: PRT, 711 AA.
ID_HFFL_HUMAN 013350; Q14870;
AC P26927; 013350; Q14870;
DT 01-AUG-1992 (Rel. 23, Created)
DT 01-AUG-1992 (Rel. 23, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Hepatocyte growth factor-like protein precursor (Macrophage
DE stimulatory protein) (MSP) (Macrophage stimulating protein).
CN MST1 OR HGFL.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_Taxid=9606;
[1]
RN SEQUENCE FROM N.A.
RC TISSUE=Liver;
RX MEDLINE=9202016; PubMed=1655021;
RA Han S., Stuart L.A., Friesner Degen S.J.;
RT "Characterization of the DNFI52 locus on human chromosome 3:
RT identification of a gene coding for four kringle domains with
RT homology to hepatocyte growth factor."
RL Biochemistry 30:9768-9780(1991).
[2]
RN SEQUENCE FROM N.A.
RC TISSUE=Liver;
RX MEDLINE=93340141; PubMed=8393443;
RA Yoshimura T., Yuhki N., Wang M.H., Skeel A., Leonard E.J.;
RT "Cloning, sequencing, and expression of human macrophage stimulating
RT protein (MSP, MST1) confirms MSP as a member of the family of kringle
RL proteins and locates the MSP gene on chromosome 3."
RL J. Biol. Chem. 268:15461-15468(1993).
CC -1- FUNCTION: PROBABLY HAS NO PROTEOLYTIC ACTIVITY, SINCE CRUCIAL AA
CC CHARACTERISTIC OF SERINE PROTEASES CATALYTIC SITES ARE NOT
CC CONSERVED.
CC -1- PTM: MAY BE CLEAVED AFTER AA 484, TO YIELD A TWO-CHAIN MOLECULE
CC HELD TOGETHER BY DISULFIDE BONDS, OR TWO SEPARATE POLYPEPTIDES.
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1. PLASMINOGEN SUBFAMILY.
CC -1- SIMILARITY: CONTAINS 4 KRINGLE DOMAINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M74178; AAA50165.1; -
DR EMBL; U37055; AAC50471.1; -
DR EMBL; L11924; AAA59872.1; -
DR PIR; A40331; A40331.
DR HSSP; P00747; 2PK4.
DR MEROPS; S01.975; -.
DR Genew; HGNC:7380; MST1.
DR MIM; 142408; -.
DR InterPro; IPR001314; Chymotrypsin.
DR InterPro; IPR000001; Kringle.
DR InterPro; IPR003609; Pan_app.
DR InterPro; IPR001254; Ser_protease_Try.
DR Pfam; PF00024; PAN; 1.
DR Pfam; PF00051; kringle; 4.
DR Pfam; PF00089; trypsin; 1.
DR PRINTS; PR00722; CHYMOTRYPSIN.
DR PRINTS; PR00018; KRINGLE.
DR ProDom; PD000395; Kringle; 4.

```

```

DR SMART; SM00130; KR; 4.
DR SMART; SM00473; PAN_AP; 1.
DR SMART; SM00020; Tryp_Spec; 1.
DR PROSITE; PS00021; KRINGLE_1; 4.
DR PROSITE; PS00070; KRINGLE_2; 4.
DR PROSITE; PS0240; TRYPSIN_DOM; 1.
DR Kringle; Glycoprotein; Serine protease homolog; Repeat; Signal;
  Polymorphism.
KM SIGNAL 1 18
FT CHAIN 19 711 HEPATOCYTE GROWTH FACTOR-LIKE PROTEIN.
FT DOMAIN 32 109 PAP.
FT DOMAIN 110 186 KRINGLE 1.
FT DOMAIN 191 268 KRINGLE 2.
FT DOMAIN 283 361 KRINGLE 3.
FT DOMAIN 370 448 KRINGLE 4.
FT DOMAIN 484 711 SERINE PROTEASE-LIKE.
FT DISULFID 56 78
FT DISULFID 60 66 BY SIMILARITY.
FT DISULFID 110 186 BY SIMILARITY.
FT DISULFID 131 169 BY SIMILARITY.
FT DISULFID 157 181 BY SIMILARITY.
FT DISULFID 191 268 BY SIMILARITY.
FT DISULFID 194 324 INTERCHAIN (BY SIMILARITY).
FT DISULFID 212 251 BY SIMILARITY.
FT DISULFID 240 263 BY SIMILARITY.
FT DISULFID 283 361 BY SIMILARITY.
FT DISULFID 304 343 BY SIMILARITY.
FT DISULFID 332 355 BY SIMILARITY.
FT DISULFID 370 448 BY SIMILARITY.
FT DISULFID 391 431 BY SIMILARITY.
FT DISULFID 419 443 BY SIMILARITY.
FT DISULFID 468 588 INTERCHAIN (BY SIMILARITY).
FT DISULFID 507 523 BY SIMILARITY.
FT DISULFID 602 667 BY SIMILARITY.
FT DISULFID 632 645 BY SIMILARITY.
FT DISULFID 657 685 BY SIMILARITY.
FT CARBOHYD 72 72 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 296 296 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 615 615 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT VARIANT 13 13 Y -> C.
FT VARIANT 212 212 C -> F.
FT VARIANT 212 212 /FTID=VAR_006631.
FT CONFLICT 623 623 L -> F (IN REF. 2).
FT SEQUENCE 711 AA; 80379 MW; 596ED21F180290E4 CRC64;

Query Match 14.5%; Score 80; DB 1; Length 711;
Best Local Similarity 26.9%; Pred. No. 0.44;
Matches 32; Conservative 17; Mismatches 30; Indels 40; Gaps 6;

OY 4 LTLRNGDQGGI-----HQP-SGVKVPAGYM-----PSGM 32
DB 496 VLRN-RGGHFCCGSLVKEQWILFARCFSSCHMPLTGYEWLGLTQNPQHGPRL- 552
OY 33 SARORRPAQLT---SGFDIAIVMLAQWNLQSGIRVISLPPSPDIPPGTGVFIYVG 88
DB 553 ---GRVPVAKWVCPSGSGQLVLTKLERSVTLNQRVALLICLPPEWVYVPPGTKEIAGWG 608

RESULT 27
EL1_HUMAN STANDARD; PRT; 258 AA.
AC Q9UN11;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Elastase 1 precursor (EC 3.4.21.36).
GN ELA1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OK NCBI_TaxId=9606;
RN [1]

```

```

RP SEQUENCE FROM N.A.
RX MEDLINE=20087075; Pubmed=10620133;
RA Tades U., Dunlop J., Khalaf S., Leigh I.M., Kelsell D.P.;
RT "Human elastase 1: evidence for expression in the skin and the
  identification of a frequent frameshift polymorphism.";
RL J. Invest. Dermatol. 114:165-170(2000).
CC -1- FUNCTION: ACTS UPON ELASTIN.
CC -1- CATALYTIC ACTIVITY: Hydrolysis of proteins, including elastin.
CC -1- SUBCELLULAR LOCATION: Secreted.
CC -1- TISSUE SPECIFICITY: Skin; not expressed in pancreas.
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1. ELASTASE SUBFAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
  between the Swiss Institute of Bioinformatics and the EMBL outstation -
  the European Bioinformatics Institute. There are no restrictions on its
  use by non-profit institutions as long as its content is in no way
  modified and this statement is not removed. Usage by and for commercial
  entities requires a license agreement (See http://www.isb-sib.ch/announce/
  or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AF120493; AAD28441.1; -.
DR HSSP; P00772; IEA1.
DR MEROPS; S01.153; -.
DR MIM; 130120; -.
DR InterPro; IPR001314; Chymotrypsin.
DR InterPro; IPR001254; Ser_protease_Try.
DR Pfam; PF00089; trypsin_1.
DR PRINTS; PR00722; CHYMOTRYPSIN.
DR SMART; SM00020; TRYPSIN_DOM; 1.
DR PROSITE; PS00134; TRYPSIN_HIS; 1.
DR PROSITE; PS00135; TRYPSIN_SER; 1.
KW Hydrolyase; Serine protease; Zymogen; Signal.
FT SIGNAL 1 8 BY SIMILARITY.
FT PROPEP 9 18 ACTIVATION PEPTIDE.
FT CHAIN 19 258 ELASTASE 1.
FT ACT_SITE 63 63 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE 111 111 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE 206 206 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT DISULFID 48 64 BY SIMILARITY.
FT DISULFID 145 212 BY SIMILARITY.
FT DISULFID 176 192 BY SIMILARITY.
FT DISULFID 202 232 BY SIMILARITY.
FT CARBOHYD 79 79 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 233 233 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 258 AA; 27812 MW; 0B9CDB8F1F010B32 CRC64;

Query Match 14.2%; Score 78.5; DB 1; Length 258;
Best Local Similarity 33.3%; Pred. No. 0.2;
Matches 19; Conservative 13; Mismatches 18; Indels 7; Gaps 1;

OY 45 SGFDIAIVMLAQWNLQSGIRVISLPPSPDIPPGTGVFIYVGDDNDSDRSKRG 101
DB 108 AGYDIALRLASQVTLNYSVQLGVLPGCAIILANNSPCVITGWR-----TKING 157

RESULT 28
TRYG_HUMAN STANDARD; PRT; 321 AA.
AC Q9NRR2; Q9NRR8; Q9C015; Q9UBR2;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Trypsin gamma precursor (EC 3.4.21.-) (Transmembrane tryptase).
GN TP5G1 OR TMT.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OK NCBI_TaxId=9606;
RN [1]
SEQUENCE FROM N.A. (VARIANTS GAMMA-1 AND GAMMA-2).

```

RX MEDLINE=20302813; PubMed=10843716;  
 RA Caughey G.H., Raymond W.W., Blount J.L., Hau L.W., Pallaro M.,  
 RA Wolters P.J., Varghese G.M.;  
 RT "Characterization of human gamma-tryptases, novel members of the  
 RT chromosome 16p mast cell tryptase and prostasin gene families.";   
 RL J. Immunol. 164:6566-6575(2000).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=99452974; PubMed=10521469;  
 RA Wong G.W., Tang Y., Feyfante E., Sali A., Li L., Li Y., Huang C.,  
 RA Friend D.S., Krilis S.A., Stevens R.L.;  
 RT "Identification of a new member of the tryptase family of mouse and  
 RT human mast cell proteases which possesses a novel COOH-terminal  
 RT hydrophobic extension.";   
 RL J. Biol. Chem. 274:30784-30793 (1999).  
 RN [3]  
 RP SEQUENCE OF 220-321 FROM N.A.  
 RA Mitten S., Agnew W.S.;  
 RT "Organization and alternative splicing of CACNA1H,";  
 RL Submitted (JAN-2001) to the EMBL/GenBank/DBJ databases.  
 CC -1- SUBCELLULAR LOCATION: MEMBRANE-ANCHORED (Potential).  
 CC -1- TISSUE SPECIFICITY: Expressed in many tissues.  
 CC -1- POLYMORPHISM: There are two alleles; gamma-I and gamma-II which  
 CC differ by 5 residues.  
 CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1. TRYPTASE SUBFAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.ebi.ac.uk/announcements>  
 CC or send an email to [license@ebi.ac.uk](mailto:license@ebi.ac.uk)).  
 CC -----  
 DR EMBL: AF191031; AAF6457.1; -  
 DR EMBL: AF195508; AAF76458.1; -  
 DR EMBL: AF175759; AAF03697.1; -  
 DR EMBL: AF175522; AAF03695.1; -  
 DR EMBL: AF223563; AAG4852.2; -  
 DR HSSP: P00763; IDPO.  
 DR MEROPS: S01.0281; -  
 DR Genew: HGNC:14134; TPST1.  
 DR InterPro: IPR001314; Chymotrypsin.  
 DR InterPro: IPR001254; Ser protease\_Try.  
 DR Pfam: PF00089; trypsin\_1.  
 DR PRINTS: PR00722; CHYMOTRPSIN.  
 DR SMART: SM00020; TRYP\_SPC; 1.  
 DR PROSITE: PS50240; TRYP\_SIN\_DOM; 1.  
 DR PROSITE: PS00134; TRYP\_SIN\_HIS; 1.  
 DR PROSITE: PS00135; TRYP\_SIN\_SER; FALSE NEG.  
 KM Hydrolyase; Serine protease; Signal; Glycoprotein; Zymogen;  
 KW Transmembrane; Polymorphism.  
 FT SIGNAL 1 19  
 FT CHAIN 20 36 TRYPTASE GAMMA LIGHT CHAIN.  
 FT CHAIN 38 321 TRYPTASE GAMMA HEAVY CHAIN.  
 FT TRANSMEM 284 304 POTENTIAL.  
 FT ACT\_SITE 78 78 CHARGE RELAY SYSTEM (BY SIMILARITY).  
 FT ACT\_SITE 125 125 CHARGE RELAY SYSTEM (BY SIMILARITY).  
 FT ACT\_SITE 222 222 CHARGE RELAY SYSTEM (BY SIMILARITY).  
 FT DISULFID 26 145 INTERCHAIN (POTENTIAL).  
 FT DISULFID 63 79 BY SIMILARITY.  
 FT DISULFID 159 228 BY SIMILARITY.  
 FT DISULFID 192 210 BY SIMILARITY.  
 FT DISULFID 218 246 BY SIMILARITY.  
 FT CARBOHYD 85 85 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT VARIANT 60 60 M -> V (IN GAMMA-II).  
 FT VARIANT 126 126 /FtId=VAR\_012097.  
 FT VARIANT 132 132 I -> M (IN GAMMA-II).  
 FT VARIANT 132 132 /FtId=VAR\_012098.  
 FT VARIANT 132 132 S -> T (IN GAMMA-II).  
 FT VARIANT 204 204 /FtId=VAR\_012099.  
 FT VARIANT 204 204 L -> I (IN GAMMA-II).  
 FT VARIANT 204 204 /FtId=VAR\_012100.

FT VARIANT 288 288 L -> F (IN GAMMA-II).  
 FT CONFLICT 160 160 /FtId=VAR\_012101.  
 FT CONFLICT 160 160 W -> S (IN REF. 1).  
 SQ SEQUENCE 321 AA; 33827 MW; FFF7B06E3C4A962D CRC64;  
 Query Match 14.1%; Score 78; DB 1; Length 321;  
 Best Local Similarity 33.7%; Pred. NO. 0.3;  
 Matches 29; Conservative 9; Mismatches 24; Indels 24; Gaps 4;  
 QY 3 FLTUNGGQGGHHQPSGVKVA PGWPMSCMSARQRPIAQTLSGPDIAVMAQWNIQS 62  
 DB 104 FSTV---QIIHSSPSC---QPG-----TSG-DIALVELSVPTLSS 139  
 QY 63 GIRVSLPQSPDIPPPGTGVIVGYG 88  
 DB 140 RILPVCLPEASDPFCRGRCWVTGWG 165  
 RESULT 29  
 CAC3 BOVIN STANDARD; PRT; 253 AA.  
 ID CAC3 BOVIN  
 AC P05805;  
 DT 01-NOV-1988 (Rel. 09, Created)  
 DT 15-DEC-1998 (Rel. 37, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE Procarboxypeptidase B precursor (Procarboxypeptidase A complex component  
 DE III) (Procarboxypeptidase A-S6 subunit III) (PROCBA-S6 III).  
 OS Bos taurus (Bovine).  
 OC Mammalia; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Eumetazoa; Euteleostomi; Cetartiodactyla; Ruminantia; Pecora; Bovidae;  
 OC Bovidae; Bovinae; Bos.  
 OX NCBI\_TaxID=9913;  
 RN [1]  
 RP SEQUENCE OF 1-25.  
 RX MEDLINE=91095520; PubMed=2269366;  
 RA Paschal R., Vendrell J., Aviles F.X., Bonicel J., Wicker C.,  
 RA Puigserver A.;  
 RT "Autolysis of procarboxypeptidase B in bovine procarboxypeptidase A ternary  
 RT complex gives rise to subunit III";  
 RL FEBS Lett. 277:37-41(1990).  
 RN [2]  
 RP SEQUENCE OF 14-253, AND DISULFIDE BONDS.  
 RX MEDLINE=86220198; PubMed=3519215;  
 RA Venot N., Sciaky M., Puigserver A., Desnuelle P., Laurent G.;  
 RT "Amino acid sequence and disulfide bridges of subunit III, a  
 RT defective endopeptidase present in the bovine pancreatic 6 S  
 RT procarboxypeptidase A complex.";   
 RL Eur. J. Biochem. 157:91-99(1986).  
 RN [3]  
 RP X-RAY CRYSTALLOGRAPHY (1.7 ANGSTROMS).  
 RX MEDLINE=94222022; PubMed=8164476;  
 RA Pignol D., Gaboriand C., Michon T., Kerfelec B., Chapus C.,  
 RA Fontecilla-Camps J.C.;  
 RT "Crystal structure of bovine procarboxypeptidase A-S6 subunit III, a  
 RT highly structured truncated zymogen B.";   
 RL EMBJ J. 13:1763-1771(1994).  
 CC -1- FUNCTION: DEFECTIVE ELASTASE-LIKE SERINE PROTEASE. DOES NOT SEEM  
 CC TO HAVE A PROTEASE ACTIVITY. ITS LIKELY FUNCTION IS TO PROTECT  
 CC PROCARBOXYPEPTIDASE A AGAINST DENATURATION IN THE ACIDIC  
 CC ENVIRONMENT OF THE RUMINANT DUODENUM.  
 CC -1- SUBUNIT: HETERODIMER OF SUBUNIT III; CARBOXYPEPTIDASE A AND  
 CC CHYMOTRYPSINOGEN C.  
 CC -1- SUBCELLULAR LOCATION: Extracellular.  
 CC -1- TISSUE SPECIFICITY: PANCREAS.  
 CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1.  
 CC PIR: A25065; CPROA3.  
 DR PDB: 1FON; 14-OCT-96.  
 DR MEROPS: S01.983; -  
 DR InterPro: IPR001314; Chymotrypsin.  
 DR InterPro: IPR001254; Ser protease\_Try.  
 DR Pfam: PF00089; trypsin\_1.  
 DR PRINTS: PR00722; CHYMOTRPSIN.  
 DR SMART: SM00020; TRYP\_SPC; 1.

DR PROSITE; PS50240; TRYPSIN\_DOM; 1.  
 DR PROSITE; PS00134; TRYPSIN\_HIS; 1.  
 DR PROSITE; PS00135; TRYPSIN\_SER; 1.  
 KM Serine protease homolog; Pancreas; Digestion; 3D-structure.  
 FT PROPEP 1 11 ACTIVATION PEPTIDE.  
 FT CHAIN 12 253 PROPROTEINASE E.  
 FT DISULFID 41 57  
 FT DISULFID 100 103  
 FT DISULFID 140 206  
 FT DISULFID 171 187  
 FT DISULFID 196 227  
 SQ SEQUENCE 253 AA; 27337 MW; 24663724D8AE409C CRC64;  
 Query Match 14.0%; Score 77.5; DB 1; Length 253;  
 Best Local Similarity 27.2%; Pred. No. 0.26;  
 Matches 22; Conservative 15; Mismatches 33; Indels 11; Gaps 1;  
 QY 9 GDQCGIHQPGGVKAPGPMSCMSARQRRPPIAQLSGFDIAIVMLAQMNLSGIRVYS 68  
 DB 78 GSEQVTPINAGDLFVHPPLMNSVCAC-----GNDIALVTLSSAQIGDKVQLAN 126  
 QY 69 LPPSDIPPEPGVIVGYGR 89  
 DB 127 LPPAGDILPNEAPCYISGMGR 147  
 RESULT 30  
 EL2\_RAT STANDARD; PRT; 271 AA.  
 AC P00774;  
 DT 21-JUL-1986 (Rel. 01, Created)  
 DT 21-JUL-1986 (Rel. 01, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE Elastase 2 precursor (EC 3.4.21.71).  
 GN ELA2.  
 OS Rattus norvegicus (Rat).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
 OX NCBI\_TaxID=10116;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=82182967; PubMed=6918221;  
 RA McDonald R.J., Swift G.H., Quinto C., Swain W., Picet R.L.,  
 RA Nikovits W., Rutter W.J.,  
 RT "Primary structure of two distinct rat pancreatic preproelastases  
 determined by sequence analysis of the complete cloned messenger  
 ribonucleic acid sequences";  
 RT Biochemistry 21:1453-1463(1982).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=85054882; PubMed=6094548;  
 RA Swift G.H., Craik C.S., Steary S.J., Quinto C., Lahaie R.G.,  
 RA Rutter W.J., McDonald R.J.,  
 RT "Structure of the two related elastase genes expressed in the rat  
 pancreas";  
 RT J. Biol. Chem. 259:14271-14278(1984).  
 RL J. Biol. Chem. 259:14271-14278(1984).  
 CC -1- FUNCTION: ACTS UPON ELASTIN.  
 CC -1- CATALYTIC ACTIVITY: Preferential cleavage: Leu-|Xaa, Met-|Xaa  
 and Phe-|Xaa. Hydrolyzes elastin.  
 CC -1- SUBCELLULAR LOCATION: Secreted.  
 CC -1- TISSUE SPECIFICITY: PANCREAS.  
 CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1, ELASTASE SUBFAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation-  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
 CC -----  
 DR EMBL; V01233; CAA24543.1; -  
 DR EMBL; L00124; AAA98780.1; -

DR EMBL; L00118; AAA98780.1; JOINED.  
 DR EMBL; L00119; AAA98780.1; JOINED.  
 DR EMBL; L00120; AAA98780.1; JOINED.  
 DR EMBL; L00121; AAA98780.1; JOINED.  
 DR EMBL; L00122; AAA98780.1; JOINED.  
 DR EMBL; L00123; AAA98780.1; JOINED.  
 DR PIR; A00961; ELK2.  
 DR HSP; P00772; IELG.  
 DR MEROPS; S01.155; -  
 DR InterPro; IPR001314; Chymotrypsin.  
 DR InterPro; IPR001254; Ser protease\_Try.  
 DR Pfam; PF00089; trypsin\_1.  
 DR PRINTS; PR00722; CHYMOTRYPSIN.  
 DR SMART; SM00020; TRYP\_SPE; 1.  
 DR PROSITE; PS50240; TRYPSIN\_DOM; 1.  
 DR PROSITE; PS00134; TRYPSIN\_HIS; 1.  
 DR PROSITE; PS00135; TRYPSIN\_SER; 1.  
 KM Hydrolyase; Serine protease; Zymogen; Signal.  
 FT SIGNAL 1 16  
 FT PROPEP 17 30 ACTIVATION PEPTIDE.  
 FT CHAIN 31 271 ELASTASE 2.  
 FT ACT\_SITE 75 75 CHARGE RELAY SYSTEM (BY SIMILARITY).  
 FT ACT\_SITE 123 123 CHARGE RELAY SYSTEM (BY SIMILARITY).  
 FT ACT\_SITE 218 218 CHARGE RELAY SYSTEM (BY SIMILARITY).  
 FT DISULFID 60 76 BY SIMILARITY.  
 FT DISULFID 157 224 BY SIMILARITY.  
 FT DISULFID 188 204 BY SIMILARITY.  
 FT DISULFID 214 245 BY SIMILARITY.  
 SQ SEQUENCE 271 AA; 28885 MW; 125C783B857B71E3 CRC64;  
 Query Match 14.0%; Score 77.5; DB 1; Length 271;  
 Best Local Similarity 44.0%; Pred. No. 0.28;  
 Matches 22; Conservative 5; Mismatches 22; Indels 1; Gaps 1;  
 QY 41 AQTLIS-GEPIALVMLAQMNLSGIRVISTLPQPSDIPPEPGGVIVGYGR 89  
 DB 115 AQLKSNMDIALVKLASPVALTSTKIQTACLPAGTILPNNYPCVTGTGMR 164  
 RESULT 31  
 CLCR\_RAT STANDARD; PRT; 268 AA.  
 ID CLCR\_RAT  
 AC P55091; Q63188;  
 DT 01-OCT-1996 (Rel. 34, Created)  
 DT 01-OCT-1996 (Rel. 34, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Caldesin precursor (EC 3.4.21.2) (Chymotrypsin C) (Serum calcium-  
 decreasing factor).  
 GN CTDC.  
 OS Rattus norvegicus (Rat).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
 OX NCBI\_TaxID=10116;  
 RN [1]  
 RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.  
 RC TISSUE=PANCREAS;  
 RX MEDLINE=96107178; PubMed=8530454;  
 RA Tomomura A., Tomomura M., Fukushima T., Akiyama M., Kubota N.,  
 RA Kumaki K., Nishii Y., Noikura T., Saheki T.,  
 RT "Molecular cloning and expression of serum calcium-decreasing factor  
 (caldesin)";  
 RT J. Biol. Chem. 270:30315-30321(1995).  
 RL J. Biol. Chem. 270:30315-30321(1995).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=PANCREAS;  
 RX MEDLINE=92165057; PubMed=1537555;  
 RA Kang J., Wiegand U., Mueller-Hill B.,  
 RT "Identification of cDNAs encoding two novel rat pancreatic serine  
 proteases";  
 RT Gene 110:181-187(1992).  
 RN [3]  
 RP CHARACTERIZATION.  
 RX MEDLINE=98207038; PubMed=9538241;



RA Yoshino-Yasuda I., Kobayashi K., Akiyama M., Itoh H., Tomomura A.,  
 RA "Sheldi T.,"  
 RT "eladectin is a novel-type serine protease expressed in pancreas, but  
 RT its homologue, elastase IV, is an artifact during cloning derived  
 RT from caldesin gene".  
 RL J. Biochem. 123:546-554(1998).  
 CC -1- FUNCTION: HAS CHYMOTRYPSIN-TYPE PROTEASE ACTIVITY AND HYPOCALCEMIC  
 CC ACTIVITY.  
 CC -1- CATALYTIC ACTIVITY: Preferential cleavage: Leu-Xaa, Tyr-Xaa,  
 CC phe-Xaa, Met-Xaa, Trp-Xaa, Gln-Xaa, Asn-Xaa.  
 CC -1- TISSUE SPECIFICITY: PANCREAS.  
 CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1. ELASTASE SUBFAMILY.  
 CC -1- CAUTION: WAS ORIGINALLY (REF.2) THOUGHT TO BE ELASTASE IV.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL, S80379; AAB35830.1; -;  
 DR EMBL, X59014; CAA41753.1; -;  
 DR HSSP, P00766; ICHG.  
 DR MEROPS, S01.157; -;  
 DR InterPro: IPR001314; Chymotrypsin.  
 DR InterPro: IPR001254; Ser\_protease\_Try.  
 DR Pfam, PF00089; trypsin\_1.  
 DR PRINTS; PR00722; CHYMOTRYPSIN.  
 DR SMART; SM00020; TRYSP\_SPC; 1.  
 DR PROSITE; PSS0240; TRYPSIN\_DOM; 1.  
 DR PROSITE; PS00134; TRYPSIN\_HIS; 1.  
 DR PROSITE; PS00135; TRYPSIN\_SER; 1.  
 DR KME Hydroxylase; Serine protease; Glycoprotein; Zymogen; Signal.  
 FT SIGNAL 1 16  
 FT PROPEP 17 29  
 FT CHAIN 30 268  
 FT ACT\_SITE 74 74  
 FT ACT\_SITE 123 123  
 FT ACT\_SITE 216 216  
 FT DISULFID 17 141  
 FT DISULFID 59 75  
 FT DISULFID 155 222  
 FT DISULFID 186 202  
 FT DISULFID 212 243  
 FT CARBOHYD 25 25  
 FT CARBOHYD 90 90  
 FT CARBOHYD 42 42  
 FT CONFLICT 96 120  
 FT  
 SQ SEQUENCE 268 AA; 29374 MW; 33B67AF34D0F8583 CRC64;  
 Query Match 13.8%; Score 76; DB 1; Length 268;  
 Best Local Similarity 35.7%; Pred. No. 0.4;  
 Matches 15; Conservative 10; Mismatches 17; Indels 0; Gaps 0;  
 QY 48 DIAIWMIAQWVNLQSGRIVSLPQPSDIPPGTGVIVYGR 89  
 DB 121 DIAIKLAEPVELSNTIIOVACIPBEGSILPDQPCYVGMGR 162  
 RESULT 32  
 HGFL\_MOUSE STANDARD; PRT; 716 AA.  
 ID HGFL\_MOUSE  
 AC P26928;  
 DT 01-AUG-1992 (Rel. 23, Created)  
 DT 01-AUG-1992 (Rel. 23, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE Hepatocyte growth factor-like protein precursor (Macrophage  
 DE stimulatory protein) (MSP).  
 GN MST1 OR HGFL.  
 OS Mus musculus (Mouse).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OC NCBI\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=BALB/c; TISSUE=Liver;  
 RX MEDLINE=92002017; Pubmed=1812957;  
 RA Fieznier Degen S.J., Stuart L.A., Han S., Jamison C.S.;  
 RT "Characterization of the mouse cDNA and gene coding for a hepatocyte  
 RT growth factor-like protein: expression during development.";,  
 RL Biochemistry 30:9781-9791(1991).  
 CC -1- FUNCTION: PROBABLY HAS NO PROTEOLYTIC ACTIVITY, SINCE CRUCIAL AA  
 CC CHARACTERISTIC OF SERINE PROTEASES CATALYTIC SITES ARE NOT  
 CC CONSERVED.  
 CC -1- TISSUE SPECIFICITY: LIVER. LOWER LEVELS IN LUNG, PLACENTA AND  
 CC ADRENAL.  
 CC -1- DEVELOPMENTAL STAGE: IS EXPRESSED AT LOW LEVELS DURING GESTATION.  
 CC JUST BEFORE BIRTH THE LEVEL INCREASES DRAMATICALLY AND REMAINS  
 CC STABLE AFTERWARDS.  
 CC -1- PTM: MAY BE CLEAVED AFTER AA 488, TO YIELD A TWO-CHAIN MOLECULE  
 CC HELD TOGETHER BY DISULFIDE BONDS, OR CLEAVED INTO TWO SEPARATE  
 CC POLYPEPTIDES.  
 CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1. PLASMINOGEN SUBFAMILY.  
 CC -1- SIMILARITY: CONTAINS 4 KRINGLE DOMAINS.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL, M74180; AAA50166.1; -;  
 DR EMBL, M74181; AAA50167.1; -;  
 DR HSSP, P00747; IKNR.  
 DR MEROPS, S01.975; -;  
 DR MGD; MGI:96080; Hgfl.  
 DR InterPro: IPR001314; Chymotrypsin.  
 DR InterPro: IPR000001; Kringle.  
 DR InterPro: IPR003014; PAN.  
 DR InterPro: IPR003609; Pan\_app.  
 DR InterPro: IPR001254; Ser\_protease\_Try.  
 DR Pfam; PF00024; PAN; 1.  
 DR Pfam; PF00051; kringle; 4.  
 DR Pfam; PF00089; trypsin\_1.  
 DR PRINTS; PR00722; CHYMOTRYPSIN.  
 DR PRINTS; PR00018; KRINGLE.  
 DR PRODOM; PD000395; Kringle; 4.  
 DR SMART; SM00430; KR; 4.  
 DR SMART; SM00473; PAN\_AP; 1.  
 DR SMART; SM00020; TRYSP\_SPC; 1.  
 DR PROSITE; PS00021; KRINGLE\_1; 4.  
 DR PROSITE; PS00070; KRINGLE\_2; 4.  
 DR PROSITE; PSS0240; TRYPSIN\_DOM; 1.  
 DR KME Kringle; Glycoprotein; Serine protease homology; Repeat; Signal.  
 FT SIGNAL 1 18  
 FT CHAIN 19 716  
 FT DOMAIN 19 109  
 FT DOMAIN 110 186  
 FT DOMAIN 191 268  
 FT DOMAIN 292 370  
 FT DOMAIN 379 457  
 FT DOMAIN 489 716  
 FT DISULFID 56 78  
 FT DISULFID 60 66  
 FT DISULFID 110 186  
 FT DISULFID 131 169  
 FT DISULFID 157 181  
 FT DISULFID 191 268  
 FT DISULFID 194 333  
 FT DISULFID 212 251  
 FT DISULFID 240 263  
 FT  
 PAM.  
 KRINGLE 1.  
 KRINGLE 2.  
 KRINGLE 3.  
 KRINGLE 4.  
 SERINE PROTEASE-LIKE.  
 BY SIMILARITY.  
 BY SIMILARITY.  
 BY SIMILARITY.  
 BY SIMILARITY.  
 BY SIMILARITY.  
 BY SIMILARITY.  
 INTERCHAIN (BY SIMILARITY).  
 BY SIMILARITY.  
 BY SIMILARITY.

```

FT DISULFID 292 370 BY SIMILARITY.
FT DISULFID 313 352 BY SIMILARITY.
FT DISULFID 341 364 BY SIMILARITY.
FT DISULFID 379 457 BY SIMILARITY.
FT DISULFID 400 440 BY SIMILARITY.
FT DISULFID 428 452 BY SIMILARITY.
FT DISULFID 477 593 INTERCHAIN (BY SIMILARITY).
FT DISULFID 512 528 BY SIMILARITY.
FT DISULFID 607 672 BY SIMILARITY.
FT DISULFID 637 651 BY SIMILARITY.
FT DISULFID 662 690 BY SIMILARITY.
FT CARBOHYD 72 72 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 173 173 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 305 305 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 620 620 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CONFLICT 19 P -> O (IN GENOMIC SEQUENCE).
SQ SEQUENCE 716 AA; 80588 MW; BBCE02EP85213ACC CRC64;

Query Match 13.8%; Score 76; DB 1; Length 716;
Best Local Similarity 27.0%; Pred. No. 1.2;
Matches 31; Conservative 19; Mismatches 33; Indels 32; Gaps 6;

Qy 4 LTLANGQGGH-----HGQ-SGVKVGAGN--PSCMSAR-Q 36
Db 501 VSLRN--RQGGHFCGSGSLVKEQWLTARQCICWSCHPELVGYEVLWLTGINQNPQGEANLQ 558
Qy 37 RRPFAQTU---SGFDIAVLAQWNLQSGIRVLSLPQPSDIPPGTGVFIYVG 88
Db 559 RVPAAKAVCGRAGSGLVTLKIERPVIINHVALICLPREGQVVPPTGCEIAGWG 613

RESULT 33
ID ST14 MOUSE STANDARD; PRT; 855 AA.
AC P56677;
DT 15-JUL-1999 (Rel. 38, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Suppressor of tumorigenicity 14 (EC 3.4.21.-) (Epithin).
GN ST14 OR PRSS14.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OC NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C.B.17SCID; TISSUE=Thymus;
RX MEDLINE=99216440; PubMed=10199918;
RA Kim M.G., Chen C., Lyu M.S., Cho E.G., Park D., Kozak C.,
RA Schwartz R.H.;
RT "Cloning and chromosomal mapping of a gene isolated from thymic
RT epithin, containing four LDL receptor modules and two CUB domains.";
RT Immunogenetics 49:420-428(1999).
RN [2]
RP REVISIONS TO 23; 321; 325; 343; 409-410 AND C-TERMINUS.
RC STRAIN=C.B.17SCID; TISSUE=Thymus;
RA Kim M.G., Chen C., Cho E.G., Park D., Schwartz R.H.;
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC TISSUE=Breast tumor;
RA Strauberg R.;
RL Submitted (MAR-2001) to the EMBL/GenBank/DBJ databases.
CC -1- SUBCELLULAR LOCATION: Type II membrane protein (Probable).
CC -1- TISSUE SPECIFICITY: HIGHLY EXPRESSED IN INTESTINE, KIDNEY, LUNG,
CC AND THYMOS. NOT EXPRESSED IN SKELETAL MUSCLE, LIVER, HEART,
CC TESTIS, AND BRAIN.
CC -1- SIMILARITY: BELONGS TO LIPIDASE FAMILY S1.
CC -1- SIMILARITY: CONTAINS 4 LDL-RECEPTOR CLASS A DOMAINS.
CC -1- SIMILARITY: CONTAINS 2 CUB DOMAINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration

```

```

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.1sb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL; AF042822; AAD02230.3; -.
DR EMBL; BC005496; AA05496.1; -.
DR HSPB; P20231; 1AAO.
DR MEROPS; S01_302; -.
DR MGD; MGI:1338881; St14.
DR InterPro; IPR000859; CUB_domain.
DR InterPro; IPR002172; LDL_recept_A.
DR InterPro; IPR001254; Ser_protease_Try.
DR Pfam; PF00057; 1dl_recept_a; 4.
DR Pfam; PF00089; trypsin; 1.
DR Pfam; PF00431; CUB; 2.
DR SMART; SM00042; CUB; 2.
DR SMART; SM00192; LDla; 3.
DR PROSITE; PS01180; CUB; 2.
DR PROSITE; PS01209; LDLa_1; 2.
DR PROSITE; PS50068; LDLa_2; 4.
DR PROSITE; PS50240; TRYPsin_DOM; 1.
DR PROSITE; PS00134; TRYPsin_HIS; 1.
DR PROSITE; PS00135; TRYPsin_SER; 1.
DR Signal-anchor; Glycoprotein; Hydrolase; Serine protease;
DR Transmembrane; Repeat.
FT DOMAIN 1 55
FT TRANSMEM 56 76
FT DOMAIN 77 855
FT DOMAIN 214 331
FT DOMAIN 340 444
FT DOMAIN 451 488
FT DOMAIN 489 522
FT DOMAIN 523 561
FT DOMAIN 565 604
FT DOMAIN 615 854
FT ACT_SITE 656 656
FT ACT_SITE 711 711
FT ACT_SITE 805 805
FT CARBOHYD 107 107
FT CARBOHYD 302 302
FT CARBOHYD 365 365
FT CARBOHYD 421 421
FT CARBOHYD 489 489
FT CARBOHYD 772 772
SQ SEQUENCE 855 AA; 94654 MW; 4F10E84DA2146D5 CRC64;

Query Match 13.8%; Score 76; DB 1; Length 855;
Best Local Similarity 29.7%; Pred. No. 1.5;
Matches 19; Conservative 13; Mismatches 20; Indels 12; Gaps 2;

Qy 47 FDIATVLAQWNLQSGIRVLSLPQPSDIPPGTGVFIYVGGRDNDPRKNGG----- 102
Db 710 YDIALLEKSEVSTVVRPICLPDATHVPAGKAIWVGWGH-----TXEGTGAL 761
Qy 103 ILKK 106
Db 762 ILKK 765

RESULT 34
ID GRAA MOUSE STANDARD; PRT; 260 AA.
AC P11032; P15118;
DT 01-JUL-1989 (Rel. 11, Created)
DT 01-JUL-1993 (Rel. 26, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Granzyme A precursor (EC 3.4.21.78) (T cell-specific serine protease
DE 1) (TSP-1) (CTLA-3) (Fragmentin 1) (Autocrine thymic lymphoma
DE granzyme-like serine protease).

```

GN ZMA OR CTLA-3 OR MISP-1.  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
OX NCBI\_TaxID=10090;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=C57BL/6;  
RX MEDLINE=92347841; PubMed=1639378;  
RA Ebnel K., Kramer M.D., Simon M.M.;  
RT "Organization of the gene encoding the mouse T-cell-specific serine  
RT proteinase 'granzyme A.'";  
RL Genomics 13:502-508(1992).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=C57BL/6;  
RX MEDLINE=8914453; PubMed=2976140;  
RA Bogenberger J., Haas M.;  
RT "cDNA clones from autocrine thymic lymphoma cells encode two  
RT mitogenic proteins, a serine protease and a truncated T-cell receptor  
RT beta-chain.";  
RL Oncogene Res. 3:301-312(1988).  
RN [3]  
RP SEQUENCE FROM N.A. (ISOFORMS HPI AND HF2).  
RC MEDLINE=93094270; PubMed=1460043;  
RX Herberberger R.U., Gerhenfeld H.K., Weisman I.L., Su L.;  
RT "Genomic organization of the mouse granzyme A gene. Two mRNAs encode  
RT the same mature granzyme A with different leader peptides.";  
RL J. Biol. Chem. 267:25488-25493(1992).  
RN [4]  
RP SEQUENCE OF 12-260 FROM N.A.  
RC MEDLINE=86208119; PubMed=2422755;  
RX Gerhenfeld H.K., Weisman I.L.;  
RT "Cloning of a cDNA for a T cell-specific serine protease from a  
RT cytotoxic T lymphocyte.";  
RL Science 232:854-858(1986).  
RN [5]  
RP SEQUENCE OF 29-48.  
RC MEDLINE=87215932; PubMed=3555842;  
RX Maeson D., Tschopp J.;  
RT "A family of serine esterases in lytic granules of cytolytic T  
RT lymphocytes.";  
RL Cell 49:679-685(1987).  
RN [6]  
RP SEQUENCE OF 29-53.  
RC MEDLINE=87030960; PubMed=3533635;  
RX Maeson D., Zama M., Tschopp J.;  
RT "Identification of granzyme A isolated from cytotoxic T-lymphocyte-  
RT granules as one of the proteases encoded by CTL-specific genes.";  
RL FEBS Lett. 208:84-88(1986).  
RN [7]  
RP SEQUENCE OF 29-46 FROM N.A.  
RC MEDLINE=88255076; PubMed=3260181;  
RX Simon H.G., Fruth U., Eckerskorn C., Lottepeich F., Kramer M.D.,  
RA Netz G., Simon M.M.;  
RT "Induction of T cell serine proteinase 1 (TSP-1)-specific mRNA in  
RT mouse T lymphocytes.";  
RL Eur. J. Immunol. 18:855-861(1988).  
RN [8]  
RP SEQUENCE OF 1-37 FROM N.A.  
RC MEDLINE=88272336; PubMed=3292396;  
RA Jeanne D.B., Tschopp J.;  
RT "Granzymes, a family of serine proteases released from granules of  
RT cytolytic T lymphocytes upon T cell receptor stimulation.";  
RL Immunol. Rev. 103:53-71(1988).  
RN [9]  
RP FUNCTION: THIS ENZYME IS NECESSARY FOR TARGET CELL LYSIS IN CELL-  
RP MEDIATED IMMUNE RESPONSES. IT CLEAVES AFTER LYS OR ARG. MAY BE  
RP INVOLVED IN APOPTOSIS.  
CC CATALYTIC ACTIVITY: Hydrolysis of proteins, including fibronectin  
CC type IV collagen and nucleolin. Preferential cleavage: Arg-|-Xaa,  
CC Lys-|-Xaa >> Phe-|-Xaa in small molecule substrates.  
CC -1- SUBUNIT: MONOMER. DISULFIDE-LINKED.  
CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC GRANULES OF CYTOLYTIC

```

CC CC T-LYMPHOCYTES.
CC CC -1- ALTERNATIVE PRODUCTS: 2 isoforms, HP1 (shown here) and HP2: are
CC CC produced by alternative splicing. HP1 is more abundant in lymphoid
CC CC tissues than HP2.
CC CC -1- TISSUE SPECIFICITY: FOUND IN CYTOTOXIC LYMPHOCYTES AND IN NORMAL
CC CC LYMPHOID TISSUES SUCH AS THYMUS AND SPLEEN.
CC CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1, GRAZYME SUBFAMILY.
CC CC -1- CAUTION: THE PREDICTED CLEAVAGE SITE FOR THE ACTIVATION PEPTIDE
CC CC OF HP2 IS UNCERTAIN. IT COULD HAVE EITHER 2 (ER) OR 7 (KRGCCR)
CC CC AA.
CC CC -----
CC CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC CC the European Bioinformatics Institute. There are no restrictions on its
CC CC use by non-profit institutions as long as its content is in no way
CC CC modified and this statement is not removed. Usage by and for commercial
CC CC entities requires a license agreement (See http://www.isb.ch/announce/isb-sib.ch).
CC CC or send an email to license@isb-sib.ch).
CC CC -----
CC CC EMBL; X14799; CAA32905.1; -.
CC CC EMBL; X62542; CAA44426.1; JOINED.
CC CC EMBL; X62543; CAA44426.1; JOINED.
CC CC EMBL; X60310; CAA44426.1; JOINED.
CC CC EMBL; X60311; CAA44426.1; JOINED.
CC CC EMBL; M13226; AAA40134.1; -.
CC CC EMBL; L01429; AAA99898.1; -.
CC CC EMBL; L01429; AAA99898.1; JOINED.
CC CC EMBL; L01426; AAA99898.1; JOINED.
CC CC EMBL; L01427; AAA99898.1; JOINED.
CC CC EMBL; L01441; AAA99898.1; JOINED.
CC CC EMBL; L01426; AAA99897.1; JOINED.
CC CC EMBL; L01427; AAA99897.1; JOINED.
CC CC EMBL; L01441; AAA99897.1; JOINED.
CC CC EMBL; M26183; AAA37735.1; -.
CC CC PIR; A26944; A26944.
CC CC PIR; A27640; A27640.
CC CC PIR; S10085; S10085.
CC CC PIR; S26184; S26184.
CC CC PIR; A45061; A45061.
CC CC HSSP; P12544; HP1.
CC CC MEROPS; S01.135; -.
CC CC MGD; MGI:109266; Grna.
CC CC InterPro; IPR001254; Ser_Protease_Try.
CC CC PIR; PF00089; trypsin; 1.
CC CC SMART; SM00020; TRYP_SPC; 1.
CC CC PROSITE; PSS0240; TRYP_SIN_DOM; 1.
CC CC PROSITE; PSS00134; TRYP_SIN_HIS; 1.
CC CC PROSITE; PSS00135; TRYP_SIN_SER; 1.
CC CC K01 Hydroxylase; Serine protease; Zymogen; Signal; T-cell; Cytolysis;
CC CC K01 Apoptosis; Alternative splicing.
CC CC FT SIGNAL 1 26
CC CC FT PROPEP 27 28
CC CC FT CHAIN 29 260
CC CC FT ACT_SITE 69 69
CC CC FT ACT_SITE 113 113
CC CC FT ACT_SITE 211 211
CC CC FT DISULFID 54 70
CC CC FT DISULFID 147 217
CC CC FT DISULFID 178 196
CC CC FT DISULFID 207 232
CC CC FT CARBOHYD 157 157
CC CC FT CARBOHYD 169 169
CC CC FT VARSPLIC 1 23
CC CC SO SEQUENCE 260 AA; 2859 MW; 7705352F08DEBDC64;
CC CC
CC CC Query Match 13.6%; Score 75; DB 1; Length 260;
CC CC Best Local Similarity 30.0%; Pred. No. 0.5;
CC CC Matches 15; Conservative 10; Mismatches 25; Indels 0; Gaps 0;

```



DR PROSITE; PS50240; TRYPSIN\_DOM; 1.  
 DR PROSITE; PS00134; TRYPSIN\_HIS; 1.  
 DR PROSITE; PS00135; TRYPSIN\_SER; 1.  
 KW Hydrolase; Serine protease; Zymogen; Signal; Multigene family.  
 FT SIGNAL 1 18  
 FT PROPEP 19 20  
 FT CHAIN 21 246  
 FT ACT\_SITE 65 65  
 FT ACT\_SITE 109 109  
 FT ACT\_SITE 202 202  
 FT DISULFID 50 66  
 FT DISULFID 143 208  
 FT DISULFID 174 187  
 SQ SEQUENCE 246 AA; 26813 MW; 3AC1452E064179FE CRC64;  
 Query Match 13.5%; Score 74.5; DB 1; Length 246;  
 Best Local Similarity 28.8%; Pred. No. 0.53; Indels 9; Gaps 1;  
 Matches 19; Conservative 13; Mismatches 25; Indels 9; Gaps 1;  
 QY 33 SARQRRIAGTLG-----FDIAIVLAQMVNLQSGIRVLSLPQSDIPPGTGVF 83  
 DB 85 STQKIVAKQIAHPYSFYFNLMIDMLKTKQKAKVAVDVLSLPSDFPFGKVC 144  
 QY 84 IVGYGR 89  
 DB 145 AAGWGR 150  
 RESULT 37  
 MCTL\_MOUSE  
 ID MCTL\_MOUSE STANDARD; PRT; 246 AA.  
 AC P11034;  
 DT 01-JUL-1989 (Rel. 11, Created)  
 DT 01-DEC-1992 (Rel. 24, Last sequence update)  
 DT 15-JUL-1998 (Rel. 36, Last annotation update)  
 DT Mast cell protease 1 precursor (EC 3.4.21.-) (MMCP-1).  
 GN MCP1.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OC NCBI\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=9238686; PubMed=1517575;  
 RA Gillival N., McNeil H.P., Stechechulte S., Austen K.F.,  
 RA Silberstein D., Gurish M.F., Somerville L.U., Stevens R.L.;  
 RT "IL-10 induces transcription of the gene for mouse mast cell  
 RT protease-1, a serine protease preferentially expressed in mucosal  
 RT mast cells of Trichinella spiralis-infected mice.";  
 RL J. Immunol. 149:2123-2129(1992).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=BALB/c; TISSUE=Liver;  
 RX MEDLINE=91285010; PubMed=2060576;  
 RA Huang R., Blom T., Hellman L.;  
 RT "Cloning and structural analysis of MMCP-1, MMCP-4 and MMCP-5, three  
 RT mouse mast cell-specific serine proteases.";  
 RL Eur. J. Immunol. 21:1611-1621(1991).  
 RN [3]  
 RP SEQUENCE OF 21-246.  
 RX MEDLINE=89207558; PubMed=2706264;  
 RA Le Trong H., Newlands G.F.J., Miller H.R.P., Charbonneau H.,  
 RA Neuzath H., Woodbury R.G.;  
 RT "Amino acid sequence of a mouse mucosal mast cell protease.";  
 RL Biochemistry 28:391-395(1989).  
 RN [4]  
 RP SEQUENCE OF 21-49.  
 RX MEDLINE=9371351; PubMed=8363563;  
 RA Newlands G.F.J., Knox D.P., Pirie-Shepherd S.R., Miller H.R.P.;  
 RT "Biochemical and immunological characterization of multiple  
 RT glycoforms of mouse mast cell protease 1: comparison with an isolated  
 RT murine serosal mast cell protease (MMCP-4).";  
 RL Biochem. J. 294:127-135(1993).  
 RN [2]

CC -1- FUNCTION: HAS A CHYMOTRYPSIN-LIKE ACTIVITY.  
 CC -1- SUBCELLULAR LOCATION: Secretory granules.  
 CC -1- TISSUE SPECIFICITY: MUCOSAL MAST CELLS.  
 CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1. GRANZYME SUBFAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.ebi.ch/announce/>  
 CC or send an email to [license@ebi-sib.ch](mailto:license@ebi-sib.ch)).  
 CC -----  
 DR EMBL; S44609; AAB23194.1; -.  
 DR EMBL; X68803; CA48703.1; -.  
 DR PIR; A30105; A30105.  
 DR PIR; S26041; S26041.  
 DR PIR; A46504; A46504.  
 DR HSP; P00770; 3RP2.  
 DR MEROPS; S01.141; -.  
 DR MGD; MGI:96937; MCP1.  
 DR InterPro; IPR001314; Chymotrypsin.  
 DR InterPro; IPR001254; Ser\_Protease\_Try.  
 DR Pfam; PF00089; Trypsin; 1.  
 DR PRINTS; PR00722; CHYMOTRYPSIN.  
 DR SMART; SM00020; TRYP\_SPC; 1.  
 DR PROSITE; PS50240; TRYPSIN\_DOM; 1.  
 DR PROSITE; PS00134; TRYPSIN\_HIS; 1.  
 DR PROSITE; PS00135; TRYPSIN\_SER; 1.  
 KW Hydrolase; Serine protease; Glycoprotein; Zymogen; Signal.  
 FT SIGNAL 1 18  
 FT PROPEP 19 20  
 FT CHAIN 21 246  
 FT ACT\_SITE 65 65  
 FT ACT\_SITE 109 109  
 FT ACT\_SITE 202 202  
 FT DISULFID 50 66  
 FT DISULFID 143 208  
 FT DISULFID 174 187  
 FT CARBOHYD 102 102  
 SQ SEQUENCE 246 AA; 27013 MW; 26E112B8C580154B CRC64;  
 Query Match 13.5%; Score 74.5; DB 1; Length 246;  
 Best Local Similarity 31.4%; Pred. No. 0.53; Indels 1; Gaps 1;  
 Matches 16; Conservative 12; Mismatches 22; Indels 1; Gaps 1;  
 QY 47 FDIAIVLAQMVNLQSGIRVLSLPQSDIPPGTGVFVGYGRDNDPDS 97  
 DB 108 YDIMLKLEKAEELTPVDVILPSPDFIDPGKVCWTAGWK-TGEKEPT 157  
 RESULT 38  
 EL1\_PIG  
 ID EL1\_PIG STANDARD; PRT; 266 AA.  
 AC P00772; Q29625;  
 DT 21-JUL-1986 (Rel. 01, Created)  
 DT 01-APR-1988 (Rel. 07, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE Elastase 1 precursor (EC 3.4.21.36).  
 GN ELA1.  
 OS Sus scrofa (Pig).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.  
 OC NCBI\_TaxID=9823;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=86304235; PubMed=3528137;  
 RA Shihara Y., Yoshida H., Mikayama T., Matsuki S., Tanaka J.,  
 RA Ikemura H.;  
 RT "Isolation and expression in Escherichia coli of a cDNA clone  
 RT encoding porcine pancreatic elastase.";  
 RL J. Biochem. 99:1707-1712(1986).  
 RN [2]

RP SEQUENCE FROM N.A.  
RX MEDLINE=87250343; PubMed=3648024;  
RA Tani T., Kawashima I., Furukawa H., Ohmine T., Takiguchi Y.;  
RT "Characterization of a silent gene for human pancreatic elastase I:  
RL J. Biochem. 101:591-599(1987).  
RN [3]  
RP SEQUENCE OF 27-266.  
RX MEDLINE=73229121; PubMed=4578945;  
RA Shotton D.M., Hartley B.S.;  
RT "Evidence for the amino acid sequence of porcine pancreatic  
RL elastase." J. 131:643-675(1973).  
RN [4]  
RP SEQUENCE OF 27-266.  
RX MEDLINE=70114042; PubMed=5415108;  
RA Shotton D.M., Hartley B.S.;  
RT "Amino-acid sequence of porcine pancreatic elastase and its  
RL homologues with other serine proteinases." J.  
RN Nature 225:802-806(1970).  
RN [5]  
RP X-RAY CRYSTALLOGRAPHY (2.5 ANGSTROMS).  
RX MEDLINE=78111463; PubMed=628010;  
RA Sawyer L., Shotton D.M., Campbell J.W., Wendell P.L., Muirhead H.,  
RA Watson H.C., Diamond R., Ladner R.C.;  
RT "The atomic structure of crystalline porcine pancreatic elastase at  
RL 2.5-A resolution: comparisons with the structure of  
RT alpha-chymotrypsin." J.  
RN J. Mol. Biol. 118:137-208(1978).  
RN [6]  
RP X-RAY CRYSTALLOGRAPHY (3.5 ANGSTROMS), AND ACTIVE SITE.  
RX MEDLINE=70114044; PubMed=5415110;  
RA Shotton D.M., Watson H.C.;  
RT "Three-dimensional structure of tosyl-elastase." J.  
RN Nature 225:811-816(1970).  
RN [7]  
RP X-RAY CRYSTALLOGRAPHY (2.4 ANGSTROMS) OF COMPLEX WITH INHIBITOR.  
RX MEDLINE=95006335; PubMed=7922044;  
RA Huang K., Strynadka N.C., Bernard V.D., Peanasky R.J., James M.N.;  
RT "The molecular structure of the complex of Ascaris  
RL chymotrypsin/elastase inhibitor with porcine elastase." J.  
RN Structure 2:679-689(1994).  
CC -1- FUNCTION: ACTS UPON ELASTIN.  
CC -1- CATALYTIC ACTIVITY: Hydrolysis of proteins, including elastin.  
CC -1- Preferred cleavage: Ala-Xaa.  
CC -1- SUBCELLULAR LOCATION: Secreted.  
CC -1- TISSUE SPECIFICITY: PANCREAS.  
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1. ELASTASE SUBFAMILY.  
CC -1- DATABASE: NAME=Worthington enzyme manual;  
CC WWW="http://www.worthington-biochem.com/manual/E/ES.html".  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/  
CC or send an email to license@isb-sib.ch).  
CC -----  
DR EMBL; X04036; CAA27670.1; -;  
DR EMBL; D00160; BAA0118.1; -;  
DR PIR; JS0013; ELPG.  
DR PDB; 1EST; 16-OCT-87.  
DR PDB; 2EST; 30-APR-94.  
DR PDB; 3EST; 15-OCT-92.  
DR PDB; 4EST; 15-JUL-92.  
DR PDB; 5EST; 15-APR-92.  
DR PDB; 6EST; 15-OCT-91.  
DR PDB; 7EST; 30-APR-94.  
DR PDB; 8EST; 15-JAN-93.  
DR PDB; 9EST; 31-JAN-94.  
DR PDB; 11NC; 31-JAN-94.  
DR PDB; 1JIM; 31-JAN-94.

DR PDB; 1ELA; 30-APR-94.  
DR PDB; 1ELB; 22-JUN-94.  
DR PDB; 1ELC; 30-APR-94.  
DR PDB; 1ELD; 14-FEB-95.  
DR PDB; 1ELE; 14-FEB-95.  
DR PDB; 1ELF; 10-JUL-95.  
DR PDB; 1ELG; 10-JUL-95.  
DR PDB; 1ESA; 30-APR-94.  
DR PDB; 1ESB; 30-APR-94.  
DR PDB; 1EAT; 05-APR-99.  
DR PDB; 1EAG; 07-FEB-95.  
DR PDB; 1EAT; 07-FEB-95.  
DR PDB; 1EAD; 07-FEB-95.  
DR PDB; 1BMA; 07-DEC-95.  
DR PDB; 1LVY; 27-JAN-97.  
DR PDB; 1NES; 29-JAN-96.  
DR PDB; 1BOE; 18-NOV-98.  
DR PDB; 1BTU; 16-FEB-99.  
DR MEROPS; S01.153; -;  
DR InterPro; IPR001314; Chymotrypsin.  
DR InterPro; IPR001254; Ser protease\_Try.  
DR Pfam; PF00089; trypsin; 1.  
DR PRINTS; PR00722; CHYMOTRYPSIN.  
DR SMART; SM00020; TRY\_SPC; 1.  
DR PROSITE; PS50240; TRYPsin\_DOM; 1.  
DR PROSITE; PS00134; TRYPsin\_HIS; 1.  
DR PROSITE; PS00135; TRYPsin\_SER; 1.  
KW Hydrolase; Serine protease; Zymogen; Signal; 3D-structure.  
FT SIGNAL 1 16  
FT PROPEP 17 26  
FT CHAIN 27 266  
FT ACT\_SITE 71 71  
FT ACT\_SITE 119 119  
FT ACT\_SITE 214 214  
FT DISULFID 56 72  
FT DISULFID 153 220  
FT DISULFID 184 200  
FT DISULFID 210 240  
FT CONFLICT 92 92  
FT CONFLICT 126 126  
FT CONFLICT 184 184  
FT CONFLICT 204 204  
FT STRAND 28 28  
FT TURN 29 29  
FT STRAND 31 32  
FT TURN 35 36  
FT TURN 39 40  
FT STRAND 41 48  
FT TURN 49 50  
FT STRAND 51 62  
FT TURN 63 64  
FT STRAND 65 68  
FT HELIX 70 72  
FT STRAND 79 83  
FT STRAND 87 87  
FT STRAND 96 105  
FT TURN 107 108  
FT TURN 111 112  
FT HELIX 114 116  
FT STRAND 121 125  
FT TURN 123 144  
FT TURN 149 150  
FT STRAND 153 157  
FT STRAND 160 160  
FT TURN 163 164  
FT STRAND 167 167  
FT STRAND 170 170  
FT STRAND 172 175  
FT STRAND 178 179  
FT HELIX 181 184  
FT TURN 185 185  
FT TURN 187 190  
FT HELIX 191 193  
D -> N (IN REF. 3).  
A -> G (IN REF. 2).  
C -> L (IN REF. 2).  
D -> N (IN REF. 3).  
ACTIVATION PEPTIDE.  
ELASTASE 1.  
CHARGE RELAY SYSTEM.  
CHARGE RELAY SYSTEM.  
CHARGE RELAY SYSTEM.

FT TURN 196 197  
 FT STRAND 198 201  
 FT STRAND 208 208  
 FT TURN 211 212  
 FT TURN 214 215  
 FT STRAND 217 222  
 FT TURN 223 224  
 FT STRAND 225 234  
 FT STRAND 236 236  
 FT TURN 237 238  
 FT STRAND 239 239  
 FT TURN 243 244  
 FT STRAND 247 251  
 FT TURN 252 254  
 FT TURN 255 255  
 FT HELIX 256 263  
 FT TURN 264 265  
 SQ SEQUENCE 266 AA; 28821 MW; BF07D6855BB50FE2 CRC64;

Query Match 13.5%; Score 74.5; DB 1; Length 266;  
 Best Local Similarity 33.3%; Pred. No. 0.58; Mismatches 19; Indels 7; Gaps 1;

Matches 19; Conservative 12; Mismatches 19; Indels 7; Gaps 1;  
 QY 45 SGFDIAVMTAQMVNLOSIGIRVISLPQSPDIPPGTGVFIYVGGRDDNDPFSRNG 101  
 DB 116 AGYDIALRLAQSIVTLNYSVOLGVLPFRAGTILANNSPCYITQWGL-----TRING 165

RESULT 39  
 TRN2\_MOUSE STANDARD; PRT; 246 AA.  
 ID TRN2\_MOUSE  
 AC P07146;  
 DT 01-APR-1988 (Rel. 07, Created)  
 DT 01-APR-1988 (Rel. 07, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE Trypsin II, antitonic precursor (EC 3.4.21.4) (Pretrypsinogen II).  
 GN TRY2.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 ON NCB1\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=A/J;  
 RX MEDLINE=87066713; PubMed=3641189;  
 RA Stevenson B.J., Hagenbuchle O., Wellauer P.K.;  
 RT "Sequence organisation and transcriptional regulation of the mouse  
 RT elastase II and trypsin genes.";  
 RL Nucleic Acids Res. 14:8307-8330(1986).  
 CC -1- CATALYTIC ACTIVITY: Preferential cleavage: Arg-|-Xaa, Lys-|-Xaa.  
 CC -1- SUBCELLULAR LOCATION: Extracellular.  
 CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL: X04574; CA28243.1; -;  
 DR EMBL: X04577; CA28245.1; -;  
 DR PIR: B25528; B25528.  
 DR HESP: P00763; IDPO.  
 DR MEROPS: S01.064; -;  
 DR MGD: MGI:102759; Try2.  
 DR InterPro: IPR001314; Chymotrypsin.  
 DR InterPro: IPR001254; Ser\_protease\_Try.  
 DR Pfam: PF00089; Trypsin\_1.  
 DR SMART: SM00022; CHYMOTRYPSIN.  
 DR PRINTS: PR00783; TRYPSIN\_DOM; 1.  
 DR PROSITE: PS50240; TRYPSIN\_DOM; 1.

DR PROSITE; PS00134; TRYPSIN\_HIS\_1.  
 DR PROSITE; PS00135; TRYPSIN\_SER\_1.  
 KW Hydrolyase; Serine protease; Digestion; Pancreas; Zymogen; Signal.  
 FT SIGNAL 1 15  
 FT PROPEP 16 23  
 FT CHAIN 24 246  
 FT ACT\_SITE 63 63  
 FT ACT\_SITE 107 107  
 FT ACT\_SITE 200 200  
 FT DISULFID 30 160  
 FT DISULFID 48 64  
 FT DISULFID 132 233  
 FT DISULFID 139 206  
 FT DISULFID 171 185  
 FT DISULFID 196 220  
 FT SITE 194 194  
 SQ SEQUENCE 246 AA; 26203 MW; CEF8C97AAC2D07AD CRC64;

Query Match 13.4%; Score 74; DB 1; Length 246;  
 Best Local Similarity 43.9%; Pred. No. 0.6; Mismatches 16; Indels 2; Gaps 1;

Matches 16; Conservative 5; Mismatches 16; Indels 2; Gaps 1;  
 QY 48 DIAVMTAQMVNLOSIGIRVISLPQSPDIPPGTGVFIYVG 88  
 DB 107 DIMLIKASPVTLNA--RVASVPLPSSCAPAGTCLISGWG 145

RESULT 40  
 PRN3\_HUMAN STANDARD; PRT; 256 AA.  
 ID PRN3\_HUMAN  
 AC P24158; P15637; Q9UQD8; P18078;  
 DT 01-APR-1990 (Rel. 14, Created)  
 DT 15-DEC-1998 (Rel. 37, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE Myeloblastin precursor (EC 3.4.21.76) (Leukocyte proteinase 3) (PR-3)  
 DE (PR3) (AGP7) (Wegener's autoantigen) (P29) (C-ANCA antigen)  
 DE (Neutrophil proteinase 4) (NP-4).  
 GN PRN3 OR MBN.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.  
 ON NCB1\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=92021028; PubMed=1681549;  
 RA Labbaye C., Mussette P., Cayre Y.E.;  
 RT "Wegener autoantigen and myeloblastin are encoded by a single mRNA.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 88:9253-9256(1991).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RA Lamerdin J.E., McCreedy P.M., Skowronski E., Adamson A.W.,  
 RA Burkhardt-Schultz K.J., Gordon L., Kyle A., Ramirez M., Stilwagen S.,  
 RA Phan H., Velasco N., Do L., Regala W., Terry A., Ganes J.,  
 RA Dargahan L., Poundstone P., Christensen M., Georgescu A., Avila J.,  
 RA Liu S., Attix C., Andreise T., Trankheim M., Amico-Keller G.,  
 RA Coffield J., Duarte S., Lucas S., Bruce R., Thomas P., Quan G.,  
 RA Krenmuller B., Arellano A., Montgomery M., Ow D., Nolan M., Trong S.,  
 RA Kodayoshi A., Olsen A.S., Carraro A.V.I.  
 RA Submitted (JUN-1998) to the EMBL/GenBank/DBJ databases.  
 RN [3]  
 RP SEQUENCE OF 2-256 FROM N.A., AND SEQUENCE OF 48-71 AND 156-181.  
 RX MEDLINE=91079774; PubMed=2258701;  
 RA Campanelli D., Melchior M., Fu Y., Nakata M., Shuman H., Nathan C.,  
 RA Gabay J.E.;  
 RT "Cloning of cDNA for proteinase 3: a serine protease, antidiabetic, and  
 RT autoantigen from human neutrophils.";  
 RL J. Exp. Med. 172:1709-1715(1990).  
 RN [4]  
 RP SEQUENCE OF 1-20 AND 22-256 FROM N.A.  
 RX MEDLINE=92390417; PubMed=1518849;  
 RA Zimmer M., Medcalf R.L., Fink T.M., Maltmann C., Lichter P.,  
 RA Jeanne D.E.;  
 RT "Three human elastase-like genes coordinately expressed in the

RT myelomonocyte lineage are organized as a single genetic locus on  
 RT 19pter.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 89:8215-8219(1992).  
 RN [5]  
 RP SEQUENCE OF 1-200 FROM N.A., AND VARIANT IIE-119.  
 RX MEDLINE=99123768; PubMed=9924693;  
 RA Clave E., Molldrem J., Hensel N., Raptis A., Barrett A.J.;  
 RT "Donor-recipient polymorphism of the proteinase 3 gene: A potential  
 RT target for T-cell alloresponses to myeloid leukemia.";  
 RL J. Immunother. 22:1-6(1999).  
 RN [6]  
 RP SEQUENCE OF 42-256 FROM N.A.  
 RX MEDLINE=90090622; PubMed=2598267;  
 RA Bories D., Raynal M.-C., Solomon D.H., Darzynkiewicz Z., Cayre Y.E.;  
 RT "Down-regulation of a serine protease, myeloblastin, causes growth  
 RT arrest and differentiation of promyelocytic leukemia cells.";  
 RL Cell 59:959-968(1989).  
 RN [7]  
 RP SEQUENCE OF 28-67 AND 228-244.  
 RX MEDLINE=91236723; PubMed=2033050;  
 RA Rao N.V., Wehner N.G., Marshall B.C., Gray W.R., Gray B.H.,  
 RA Hoidal J.R.;  
 RT "Characterization of proteinase-3 (PR-3), a neutrophil serine  
 RT proteinase. Structural and functional properties.";  
 RL J. Biol. Chem. 266:9540-9548(1991).  
 RN [8]  
 RP SEQUENCE OF 28-52.  
 RX MEDLINE=91025622; PubMed=2121162;  
 RA Ohlsson K., Linder C., Rosenegren M.;  
 RT "Monoclonal antibodies specific for neutrophil proteinase 4.  
 RT Production and use for isolation of the enzyme.";  
 RL Biol. Chem. Hoppe-Seyler 371:549-555(1990).  
 RN [9]  
 RP SEQUENCE OF 28-47.  
 RX MEDLINE=89315847; PubMed=2501794;  
 RA Gabay J.E., Scott R.W., Campanelli D., Griffith J., Wilde C.,  
 RA Marra M.N., Seeger M., Nathan C.F.;  
 RT "Antibiotic proteins of human polymorphonuclear leukocytes.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 86:5610-5614(1989).  
 RN [10]  
 RP SEQUENCE OF 28-47 AND 196-219.  
 RX MEDLINE=90130450; PubMed=2104977;  
 RA Wilde C.G., Shable J.L., Griffith J.E., Scott R.W.;  
 RT "Characterization of two azurophil granule proteases with active-site  
 RT homology to neutrophil elastase.";  
 RL J. Biol. Chem. 265:2038-2041(1990).  
 RN [11]  
 RP SEQUENCE OF 28-48, AND IDENTITY OF WEGENER'S AUTOANTIGEN WITH PR-3.  
 RX MEDLINE=90332035; PubMed=2377228;  
 RA Jemne D.E., Tschopp J., Luedemann J., Urecht B., Gross W.L.;  
 RT "Wegener's autoantigen decoded.";  
 RL Nature 346:520-520(1990).  
 RN [12]  
 RP IDENTITY OF WEGENER'S AUTOANTIGEN WITH PROTEINASE 3.  
 RX MEDLINE=91051123; PubMed=2242436;  
 RA Gupta S.K., Niles J.L., McCluskey R.T., Arnaout M.A.;  
 RT "Identity of Wegener's autoantigen (p29) with proteinase 3 and  
 RT myeloblastin.";  
 RL Blood 76:2162-2162(1990).  
 RN [13]  
 RP X-RAY CRYSTALLOGRAPHY (2.2 ANGSTROMS).  
 RX MEDLINE=96346178; PubMed=8757293;  
 RA Fujinaga M., Charnata M.M., Halenbeck R., Kothe K., James M.N.G.;  
 RT "The crystal structure of PR3, a neutrophil serine proteinase antigen  
 RT of Wegener's granulomatosis antibodies.";  
 RL J. Mol. Biol. 261:267-278(1996).  
 CC -1- FUNCTION: POLYMORPHONUCLEAR LEUKOCYTE SERINE PROTEASE THAT  
 CC DEGRADES ELASTIN, FIBRONECTIN, LAMININ, VITRONECTIN, AND COLLAGEN  
 CC TYPES I, III, AND IV (IN VITRO) AND CAUSES EMPHYSEMA WHEN  
 CC ADMINISTERED BY TRACHEAL INSUFFLATION TO HAMSTERS  
 CC -1- CATALYTIC ACTIVITY: Hydrolysis of proteins, including elastin, by  
 CC preferential cleavage: Ala-|-Xaa > Val-|-Xaa.  
 CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1. ELASTASE SUBFAMILY.

CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; X56132; CA39598.1; -;  
 DR EMBL; AC004799; AAC18958.1; -;  
 DR EMBL; M75154; AAB59558.1; -;  
 DR EMBL; M96839; AAB59493.1; -;  
 DR EMBL; M96838; AAB59493.1; JOINED.  
 DR EMBL; M96837; AAB59493.1; JOINED.  
 DR EMBL; X55668; CA39203.1; -;  
 DR EMBL; M29142; AAB36342.1; -;  
 DR EMBL; AF015449; AAD21524.1; -;  
 DR EMBL; AF015446; AAD21524.1; JOINED.  
 DR EMBL; AF015447; AAD21524.1; JOINED.  
 DR EMBL; AF015448; AAD21524.1; JOINED.  
 DR EMBL; M96628; AAB59364.1; -;  
 DR PIR; A43983; PRH3  
 DR PDB; 1FU7; 11-UTL-96.  
 DR MEROPS; S01.134; -;  
 DR GeneW; HGNC:9495; PRN3.  
 DR MIM; 177020; -;  
 DR InterPro; IPR001254; Ser\_protease\_Try.  
 DR InterPro; IPR001254; Ser\_protease\_Try.  
 DR Pfam; PF00089; Trypsin\_1.  
 DR PRINTS; PR00722; CHYMOTRYPSIN.  
 DR SMART; SM00020; TRY\_SPC; 1.  
 DR PROSITE; PS00240; TRYPSIN\_DOM; 1.  
 DR PROSITE; PS00134; TRYPSIN\_HIS; 1.  
 DR PROSITE; PS00135; TRYPSIN\_SER; 1.  
 KM Hydrolyase; Serine protease; Glycoprotein; Zymogen; Signal;  
 KW Polymorphism; 3D-structure.  
 FT SIGNAL 1 25  
 FT PROPEP 26 27  
 FT CHAIN 28 248  
 FT PROPEP 249 256  
 FT ACT\_SITE 71 71  
 FT ACT\_SITE 118 118  
 FT ACT\_SITE 203 203  
 FT CARBOHYD 129 129  
 FT CARBOHYD 174 174  
 FT DISULFID 56 72  
 FT DISULFID 152 209  
 FT DISULFID 182 188  
 FT DISULFID 199 224  
 FT VARIANT 119 119  
 FT VARIANT 135 135  
 FT VARIANT 136 136  
 FT VARIANT 136 136  
 FT CONFLICT 2 2  
 FT CONFLICT 46 46  
 FT CONFLICT 48 48  
 FT CONFLICT 64 64  
 FT CONFLICT 70 70  
 FT CONFLICT 255 255  
 SQ SEQUENCE 256 AA; 27807 MW; CBECA36D8C4B2A40 CRC64;  
 Query Match 13.4%; Score 74; DB 1; Length 256;  
 Best Local Similarity 31.0%; Pred. No. 0.63;  
 Matches 13; Conservative 9; Mismatches 20; Indels 0; Gaps 0;  
 Oy 48 DIATYMLAQWVNLQSGRIVISLPPQSDIPRPGTGFVIGYGR 89  
 Db 118 DVLLIQSSPNLNSAVTVOLEPQDQVPRPGTQCLAMGWR 159



Search completed: April 9, 2003, 12:56:03  
Job time : 27.1967 secs

---

**THIS PAGE BLANK (USPTO)**



















```

RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=K12;
RX MEDLINE=97061202; PubMed=8905232;
RA Oshima T., Alba H., Baba T., Fujita K., Hayashi K., Honjo A.,
RA Ikemoto K., Inada T., Itoh T., Kajihara M., Kanai K., Kashimoto K.,
RA Kimura S., Kitagawa M., Makino K., Masuda S., Miki T., Mizubuchi K.,
RA Mori H., Motomura K., Nakamura Y., Nashimoto H., Nishio Y., Saito N.,
RA Samped G., Seiki Y., Tagami H., Takemoto K., Wada C., Yamamoto Y.,
RA Yano M., Horiiuchi T.;
RT "A 718-kb DNA sequence of the Escherichia coli K-12 genome
RT corresponding to the 12.7-28.0 min region on the linkage map.";
RL DNA Res. 3:137-155(1996).
CC -1- SIMILARITY: BELONGS TO THE NNA/SNNA/SOXA/DS2A) FAMILY OF
CC MONOOXYGENASES.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb.ch/announce/
CC or send an email to license@isb.sib.ch).
CC -----
CC
CC EMBL: AE000202; AAC74097.1; ALT_INIT.
CC EMBL: D90738; BAA35789.1; ALT_INIT.
CC Ecogene: EG13859; ycdm.
CC InterPro: IPR002103; Bac_Luciferase.
CC Pfam: PF00296; bac_Luciferase.1.
CC Hypothetical protein: Oxidoreductase: Monooxygenase:
CC Complete proteome.
CC SEQUENCE 363 AA; 39943 MW; 1A35F5DC9B7B8C54 CRC64;
CC
CC Alignment Scores:
CC
CC Pred. No.: 0.714 Length: 363
CC Score: 9.00 Matches: 9
CC Percent Similarity: 100.00% Conservative: 0
CC Best Local Similarity: 100.00% Mismatches: 0
CC Query Match: 1.89% Indels: 0
CC DB: Gaps: 1 Gaps: 0
CC
CC US-10-020-441-1 (1-1477) x YCDM_ECOLI (1-363)
CC
CC QY 1430 TATGCCAGTGTGGCTAGAAATGTTGAT 1456
CC |||||||||||||||||||
CC Db 308 TYRALASERVALAALARGMETLEUASP 316
CC
CC RESULT 4
CC COAD_BUCAT STANDARD: PRT: 165 AA.
CC ID COAD_BUCAT STANDARD: PRT: 165 AA.
CC AC P57643;
CC DT 16-OCT-2001 (Rel. 40, Created)
CC DT 16-OCT-2001 (Rel. 40, Last sequence update)
CC DT 15-JUN-2002 (Rel. 41, Last annotation update)
CC DE Phosphopantetheine adenylyltransferase (EC 2.7.7.3) (Pantetheine-
CC phosphate adenylyltransferase) (PPAT) (Dephospho-CoA
CC pyrophosphorylase).
CC DE COAD OR BU583.
CC OS Buchnera aphidicola (subsp. Acyrthosiphon pisum) (Acyrthosiphon pisum
CC symbiotic bacterium).
CC OC Bacteria: Proteobacteria; gamma subdivision; Buchnera.
CC CX NCBI_TaxID=118099;
CC RM [1]
CC RP SEQUENCE FROM N.A.
CC RC STRAIN=Tokyo. 1998;
CC RX MEDLINE=20451173; PubMed=1093077;
CC RA Shigenobu S., Watanabe H., Hattori M., Sakaki Y., Ishikawa H.;
CC RT "Genome sequence of the endocellular bacterial symbiont of aphids
CC Buchnera sp. APS."
CC RT Nature 407:81-86(2000).
CC -1- FUNCTION: REVERSIBLY TRANSFERS AN ADENYLYL GROUP FROM ATP TO 4'-
CC PHOSPHOPANTHETHEINE, YIELDING DEPHOSPHO-COA (OPCOA) AND
CC PYROPHOSPHATE (BY SIMILARITY).
CC

```

```
CC -1- CATALYTIC ACTIVITY: ATP + pantotheine 4'-phosphate = diphosphate +
CC dephospho-CoA.
CC -1- PARTIAL: Coenzyme A (CoA) biosynthesis; fourth step.
CC -1- SUBUNIT: HOMOHXAMER (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
CC -1- SIMILARITY: BELONGS TO THE COAD FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: AP001119; BAB13272.1; -.
CC HSSP: P23875; 1B5T.
CC InterPro: IPR004821; Cyl_tran_rel.
CC InterPro: IPR004820; Cylidylitransf.
CC InterPro: IPR001980; LPS_biosynth.
CC Pfam: PF01467; Cylidylitransf. 1.
CC PRINTS: PRO1020; LPSBIOSYNTHS.
CC TIGRFAMs: TIGR00125; cyl_tran_rel. 1.
CC Transferrase: Nucleotidyltransferase; Coenzyme A biosynthesis;
CC Complete proteome.
CC SEQUENCE 165 AA; 18847 MW; BE8D8B49471B10B CRC64;

Alignment Scores:
Pred. No.: 9.79 Length: 165
Score: 8.00 Matches: 8
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 1.69% Indels: 0
DB: 1 Gaps: 0

US-10-020-441-1 (1-1477) x COAD_BUCAL (1-165)
QY 470 AGCATTCACATTCGATTCGAT 447
DB 31 SerileThreileAlaIleSerAsn 38

RESULT 5
YE03_ARCFU
ID YE03_ARCFU STANDARD; PRT; 219 AA.
AC 028869;
DC 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DE Hypothetical protein AF1403.
GN AF1403.
OS Archaeoglobus fulgidus.
OC Archaea; Euryarchaeota; Archaeoglobi; Archaeoglobales;
OC Archaeoglobaceae; Archaeoglobus.
OX NCBI_TaxID=2234;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=VC-16 / DSM 4304 / ATCC 49558;
RX MEDLINE=98049343; PubMed=9389475;
RA Klenk H.-P., Clayton R.A., Tomb J.-F., White O., Nelson K.E.,
RA Ketchum K.A., Dodson R.J., Gwinn M., Hickey E.K., Peterson J.D.,
RA Richardson D.L., Keriavage A.R., Graham D.E., Kyriades N.C.,
RA Fleischmann R.D., Quackenbush J., Lee N.H., Sutton G.G., Gill S.,
RA Kirtress E.F., Dougherty B.A., McKenney K., Adams M.D., Loftus B.,
RA Peterson S., Reich C.I., McNeil L.K., Badger J.H., Glodek A., Zhou L.,
RA Overbeek R., Gocayne J.D., Weidman J.F., McDonald L., Utterback T.,
RA Cotton M.D., Spriggs T., Arlisch P., Kaine B.P., Sykes S.M.,
RA Sadow P.W., D'Andrea K.P., Bowman C., Fujii C., Garland S.A.,
RA Mason T.M., Olsen G.J., Fraser C.M., Smith H.O., Woese C.R.,
RA Venter J.C.;
RT "The complete genome sequence of the hyperthermophilic, sulphate-
RT reducing archaeon Archaeoglobus fulgidus.";
RL Nature 390:364-370(1997).
CC -1- SIMILARITY: STRONG, TO M. JANNASCHII MJ1458 AND
```

```
CC M. THERMOAUTOTROPHICUM MTH1235.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: AF001007; AAB89840.1; -.
CC TIGR: AF1403; -.
CC InterPro: IPR002912; ACT.
CC Pfam: PF01842; ACT. 1.
CC Hypothetical protein; Complete proteome.
CC SEQUENCE 219 AA; 24038 MW; 35B7900C84D69263 CRC64;

Alignment Scores:
Pred. No.: 9.34 Length: 219
Score: 8.00 Matches: 8
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 1.68% Indels: 0
DB: 1 Gaps: 0

US-10-020-441-1 (1-1477) x YE03_ARCFU (1-219)
QY 493 AGTGATCAGAGTCGATCGCTG 516
DB 168 SerGlyIleArgValIleSerLeu 175

RESULT 6
MCT1_MOUSE
ID MCT1_MOUSE STANDARD; PRT; 246 AA.
AC P11034;
DC 01-JUL-1989 (Rel. 11, Created)
DT 01-DEC-1992 (Rel. 24, Last sequence update)
DT 15-JUL-1998 (Rel. 36, Last annotation update)
DE Mast cell protease 1 precursor (EC 3.4.21.-) (MMCP-1).
GN MCP1.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=9238686; PubMed=1517575;
RA Chlidyal N., McNeil H.P., Stechschulte S., Austen K.F.,
RA Silberstein D., Gurish W.F., Somerville L.L., Stevens R.L.;
RT "IL-10 induces transcription of the gene for mouse mast cell
RT protease-1, a serine protease preferentially expressed in mucosal
RT mast cells of Trichinella spiralis-infected mice.";
RL J. Immunol. 149:2123-2129(1992).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=BALB/c; TISSUE=Liver;
RX MEDLINE=91285010; PubMed=2060576;
RA Huang R., Blom T., Hellman L.;
RT "Cloning and structural analysis of MMCP-1, MMCP-4 and MMCP-5, three
RT mouse mast cell-specific serine proteases.";
RL Eur. J. Immunol. 21:1611-1621(1991).
RN [3]
RP SEQUENCE OF 21-246.
RX MEDLINE=89207558; PubMed=2706264;
RA le Trong H., Newlands G.F.J., Miller H.R.P., Charbonneau H.,
RA Neurath H., Woodbury R.G.;
RT "Amino acid sequence of a mouse mucosal mast cell protease.";
RL Biochemistry 28:391-395(1989).
RN [4]
RP SEQUENCE OF 21-49.
RX MEDLINE=93371351; PubMed=8363563;
RA Newlands G.F.J., Knox D.P., Pirie-Shepherd S.R., Miller H.R.P.;
RT "Biochemical and immunological characterization of multiple
```

RT glycoforms of mouse mast cell protease 1: comparison with an isolated  
RT murine serosal mast cell protease (MMP-4).";  
RL Biochem. J. 294:127-135(1993).  
CC -1- FUNCTION: HAS A CHYMOTRYPSIN-LIKE ACTIVITY.  
CC -1- SUBCELLULAR LOCATION: Secretory granules.  
CC -1- TISSUE SPECIFICITY: MUCOSAL MAST CELLS.  
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1. GRANZYME SUBFAMILY.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL; S44609; AAB23194.1; -;  
DR EMBL; X68803; CAA48703.1; -;  
DR PIR; A30105; A30105.  
DR PIR; S26041; S26041.  
DR PIR; A46504; A46504.  
DR HSSP; P00770; 3RP2.  
DR MEROPS; S01.141; -;  
DR MGD; MGI:96937; MCP1.  
DR InterPro; IPR001314; Chymotrypsin.  
DR InterPro; IPR001254; Ser\_protease\_Try.  
DR Pfam; PF00089; trypsin\_1.  
DR PRINTS; PR00722; CHYMOTRYPSIN.  
DR SMART; SM00020; TRYP\_SPE; 1.  
DR PROSITE; PS50240; TRYPsin\_DOM; 1.  
DR PROSITE; PS00134; TRYPsin\_HIS; 1.  
DR PROSITE; PS00135; TRYPsin\_SER; 1.  
KW Hydrolyase; Serine protease; Glycoprotein; zymogen; Signal.  
FT SIGNAL 1 18  
FT PROPEP 19 20  
FT CHAIN 21 246  
FT ACT\_SITE 65 65  
FT ACT\_SITE 109 109  
FT ACT\_SITE 202 202  
FT DISULFID 50 66  
FT DISULFID 143 208  
FT DISULFID 174 187  
FT CARBOHYD 102 102  
SQ SEQUENCE 246 AA; 27013 MW; 26E1128C580154B CRC64;  
  
Alignment Scores:  
Pred. No.: 9.16 Length: 246  
Score: 8.00 Matches: 8  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 1.68% Indels: 0  
DB: 1 Gaps: 0  
  
US-10-020-441-1 (1-1477) x MCT1\_MOUSE (1-246)  
OY 1328 GGTGACAGTGTGACACTCTCCTC 1351  
DB 200 GlyAspSerGlyGlyProLeuLeu 207  
|||||  
  
RESULT 7  
MCT2\_SHEEP  
ID MCT2\_SHEEP STANDARD; PRT; 246 AA.  
AC P79204;  
DT 30-MAY-2000 (Rel. 39, Created)  
DT 30-MAY-2000 (Rel. 39, Last sequence update)  
DE Mast cell protease 2 precursor (EC 3.4.21.-) (SMCP-2).  
OS Ovis aries (Sheep).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;  
OC Bovidae; Caprinae; Ovis.  
OX NCBI\_TaxID=9940;  
RN [1]

RP SEQUENCE FROM N.A.  
RC Tissue-Bone marrow;  
RA McAliese S.M., Knox D.P., Huntley J.F., Miller H.R.P.;  
RL Submitted (DEC-1996) to the EMBL/GenBank/DBJ databases.  
CC -1- FUNCTION: HAS A CHYMOTRYPSIN-LIKE AND TRYPsin-LIKE ACTIVITY.  
CC -1- SUBCELLULAR LOCATION: Secretory granules.  
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1. GRANZYME SUBFAMILY.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL; Y08133; CAA69327.1; -;  
DR HSSP; P23946; 1KLT.  
DR MEROPS; S01.140; -;  
DR InterPro; IPR001314; Chymotrypsin.  
DR InterPro; IPR001254; Ser\_protease\_Try.  
DR Pfam; PF00089; trypsin\_1.  
DR PRINTS; PR00722; CHYMOTRYPSIN.  
DR SMART; SM00020; TRYP\_SPE; 1.  
DR PROSITE; PS50240; TRYPsin\_DOM; 1.  
DR PROSITE; PS00134; TRYPsin\_HIS; 1.  
DR PROSITE; PS00135; TRYPsin\_SER; 1.  
KW Hydrolyase; Serine protease; zymogen; Signal.  
FT SIGNAL 1 19  
FT PROPEP 20 21  
FT CHAIN 22 246  
FT ACT\_SITE 66 66  
FT ACT\_SITE 109 109  
FT ACT\_SITE 202 202  
FT DISULFID 51 67  
FT DISULFID 143 208  
FT DISULFID 174 187  
FT CARBOHYD 120 120  
SQ SEQUENCE 246 AA; 27146 MW; 4C6AF6710B8431AD CRC64;  
  
Alignment Scores:  
Pred. No.: 9.16 Length: 246  
Score: 8.00 Matches: 8  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 1.68% Indels: 0  
DB: 1 Gaps: 0  
  
US-10-020-441-1 (1-1477) x MCT2\_SHEEP (1-246)  
OY 1328 GGTGACAGTGTGACACTCTCCTC 1351  
DB 200 GlyAspSerGlyGlyProLeuLeu 207  
|||||  
  
RESULT 8  
MCT4\_MOUSE  
ID MCT4\_MOUSE STANDARD; PRT; 246 AA.  
AC P21812;  
DT 01-MAY-1991 (Rel. 18, Created)  
DT 01-MAY-1991 (Rel. 18, Last sequence update)  
DT 15-JUL-1998 (Rel. 36, Last annotation update)  
DE Mast cell protease 4 precursor (EC 3.4.21.-) (MMP-4) (serosal mast  
DE cell protease) (MSMCP).  
GN MCP4.  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
OX NCBI\_TaxID=10090;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-BALB/c;  
RX MEDLINE=91107701; PubMed=1988455;

RA Serfin W.E., Sullivan T.P., Conder G.A., Ebrahimi A., Marcham P.,  
 RA Johnson S.S., Austen K.F., Reynolds R.S.;  
 RT "Cloning of the cDNA and gene for mouse mast cell protease 4.  
 RT Demonstration of its late transcription in mast cell subclasses and  
 RT analysis of its homology to subclass-specific neutral proteases of  
 RT the mouse and rat.";  
 RL J. Biol. Chem. 266:1934-1941(1991).  
 RN [2]  
 RP SEQUENCE OF 7-246 FROM N.A.  
 RC STRAIN=Leaden X A1; TISSUE=Connective tissue;  
 RA MEDLINE=91285010; PubMed=2060576;  
 RX Huang R., Blom T., Hellman L.;  
 RT "Cloning and structural analysis of MMCP-1, MMCP-4 and MMCP-5, three  
 RT mouse mast cell-specific serine proteases.";  
 RL Eur. J. Immunol. 21:1611-1621(1991).  
 RN [3]  
 RP SEQUENCE OF 21-43.  
 RX MEDLINE=90222202; PubMed=2326280;  
 RA Reynolds D.S., Stevens R.L., Lane W.S., Carr M.H., Austen K.F.,  
 RA Serfin W.E.;  
 RT "Different mouse mast cell populations express various combinations  
 RT of at least six distinct mast cell serine proteases.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 87:3230-3234(1990).  
 RN [4]  
 RP SEQUENCE OF 21-40.  
 RX MEDLINE=93371351; PubMed=8363563;  
 RA Newlands G.F.V., Knox D.P., Pirie-Shepherd S.R., Miller H.R.P.;  
 RT "Biochemical and immunological characterization of multiple  
 RT glycoforms of mouse mast cell protease 1: comparison with an isolated  
 RT murine serosal mast cell protease (MMCP-4).";  
 RL Biochem. J. 294:127-135(1993).  
 RN [1]  
 RP TISSUE SPECIFICITY: HAS A CHYMOTRYPSIN-LIKE ACTIVITY.  
 CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1. GRANZYME SUBFAMILY.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL: M55617; AAA39989.1; -  
 DR EMBL: M55616; AAA72939.1; -  
 DR EMBL: X68804; CAA48704.1; -  
 DR PIR: A38678; A38678.  
 DR PIR: B35646; B35646.  
 DR PIR: B38678; B38678.  
 DR HSSP: P00770; 3RP2.  
 DR MEROPS: S01.149; -  
 DR MGD: MGI:96940; Mcp4.  
 DR InterPro: IPR001314; Chymotrypsin.  
 DR Pfam: PF00089; trypsin; 1.  
 DR PRINTS: PR00722; CHYMOTRYPSIN.  
 DR SMART: SM00020; TRYP\_SPC; 1.  
 DR PROSITE: PS50240; TRTPSIN\_DOM; 1.  
 DR PROSITE: PS00134; TRYPsin\_HIS; 1.  
 DR PROSITE: PS00135; TRYPsin\_SER; 1.  
 KW Hydrolyase; Serine protease; zymogen; signal.  
 FT SIGNAL 1 18  
 FT PROPEP 19 20  
 FT CHAIN 21 246  
 FT ACT\_SITE 65 65  
 FT ACT\_SITE 109 109  
 FT ACT\_SITE 202 202  
 FT DISULFID 50 66  
 FT DISULFID 143 208  
 FT DISULFID 174 187  
 FT VARIANT 61 61  
 FT VARIANT 160 160  
 FT VARIANT 246 246  
 E -> KK (IN STRAIN LEADEN X A1).

SQ SEQUENCE 246 AA; 27203 MW; 0887A1C71ACE2698 CRC64;  
 Alignment Scores:  
 Pred. No.: 9.16 Length: 246  
 Score: 8.00 Matches: 8  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 1.68% Indels: 0  
 DB: 1 Gaps: 0  
 US-10-020-441-1 (1-1477) x MCT9\_MOUSE (1-246)  
 QY 1328 GGTCACAGTCGTGACCTCTCTC 1351  
 Db 200 GlyAsperGlyGlyProLeuLeu 207  
 ;  
 RESULT 9  
 MCT9\_MOUSE STANDARD: PRT; 246 AA.  
 ID MCT9\_MOUSE  
 AC 035164;  
 DT 15-JUL-1998 (Rel. 36, Created)  
 DT 15-JUL-1998 (Rel. 36, Last sequence update)  
 DT 15-JUL-1998 (Rel. 36, Last annotation update)  
 DE Mast cell protease 9 precursor (EC 3.4.21.-) (MMCP-9).  
 GN MCTP9.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OX NCBI\_Taxid=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=BALB/CJ.  
 RX MEDLINE=98030598; PubMed=9360993;  
 RA Hunt J.E., Friend D.S., Gurish M.F., Feyfant E., Sall A., Huang C.,  
 RA Gillyard N., Stechschulte S., Austen K.F., Stevens R.L.;  
 RT "Mouse mast cell protease 9, a novel member of the chromosome 14  
 RT family of serine proteases that is selectively expressed in uterine  
 RT mast cells.";  
 RL J. Biol. Chem. 272:29158-29166(1997).  
 CC -1- TISSUE SPECIFICITY: SELECTIVELY EXPRESSED IN UTERINE MAST CELLS.  
 CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1. GRANZYME SUBFAMILY.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL: AF007119; AAB65245.1; -  
 DR EMBL: AF007120; AAB65246.1; -  
 DR HSSP: P00770; 3RP2.  
 DR MGD: MGI:1194491; Mcp9.  
 DR InterPro: IPR001314; Chymotrypsin.  
 DR InterPro: IPR001254; Ser\_protease\_Try.  
 DR Pfam: PF00089; trypsin; 1.  
 DR PRINTS: PR00722; CHYMOTRYPSIN.  
 DR SMART: SM00020; TRYP\_SPC; 1.  
 DR PROSITE: PS50240; TRYPsin\_DOM; 1.  
 DR PROSITE: PS00134; TRYPsin\_HIS; 1.  
 DR PROSITE: PS00135; TRYPsin\_SER; 1.  
 KW Hydrolyase; Serine protease; zymogen; signal; Multigene family.  
 FT SIGNAL 1 18  
 FT PROPEP 19 20  
 FT CHAIN 21 246  
 FT ACT\_SITE 65 65  
 FT ACT\_SITE 109 109  
 FT ACT\_SITE 202 202  
 FT DISULFID 50 66  
 FT DISULFID 143 208  
 FT DISULFID 174 187  
 FT VARIANT 61 61  
 FT VARIANT 160 160  
 FT VARIANT 246 246  
 SQ SEQUENCE 246 AA; 26652 MW; C3EEFC5D7B51DD01 CRC64;

## Alignment Scores:

Pred. No.: 9.16 Length: 246  
 Score: 8.00 Matches: 8  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 1.68% Indels: 0  
 DB: 1 Gaps: 0

US-10-020-441-1 (1-1477) x MCT9\_MOUSE (1-246)

QY 1328 GGTGACAGTGTGACCTCTCCTC 1351

Db 200 GLYAspserGlyGlyProleuLeu 207

## RESULT 10

MCTX\_MOUSE STANDARD: PRT: 246 AA.

AC 000356:

DT 01-JUL-1993 (Rel. 26, Created)

DT 01-JUL-1993 (Rel. 26, Last sequence update)

DT 15-JUL-1998 (Rel. 36, Last annotation update)

DE Mast cell protease-like protein precursor (EC 3.4.21.-).

GN MCPYL.

OS Mus musculus (Mouse).

OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.

OX NCBI\_TaxID=10090;

RN [1]

RP SEQUENCE FROM N.A.

RX MEDLINE=91107701; PubMed=1988455;

RA Sealfin W.E., Sullivan T.P., Conder G.A., Edrahimi A., Marcham P.,

RA Johnson S.S., Austen K.F., Reynolds D.S.,

RT "Cloning of the CDNA and gene for mouse mast cell protease 4.

RT Demonstration of its late transcription in mast cell subclasses and

RT analysis of its homology to subclass-specific neutral proteases of

RT the mouse and rat."

RL J. Biol. Chem. 266:1941-1941(1991).

CC -1. SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1. GRANZYME SUBFAMILY.

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC -----

CC EMBL; M57401; AAA39990.1; -.

DR HSSP; P00770; 3RP2.

DR MEROPS; S01.304; -.

DR MCD; MGI:102792; McPyl.

DR InterPro; IPR001314; Chymotrypsin.

DR Pfam; PF00089; trypsin.1.

DR PRINTS; PR00722; CHYMOTRYPSIN.

DR SMART; SM00020; TRYPSIN.

DR PROSITE; PS50240; TRYPSIN\_DOM.1.

DR PROSITE; PS00134; TRYPSIN\_HIS.1.

DR PROSITE; PS00135; TRYPSIN\_SER.1.

KW HydroLase; Serine protease; Zymogen; Signal; Multigene family.

FT SIGNAL 1 18

FT PROPEP 19 20

FT CHAIN 21 246

FT ACT\_SITE 65 109

FT ACT\_SITE 109 109

FT ACT\_SITE 202 202

FT DISULFID 50 66

FT DISULFID 143 208

FT DISULFID 174 187

SO SEQUENCE 246 AA; 26774 MW; 304BA25BC2B2B2AD CRC64;

## Alignment Scores:

Pred. No.: 9.16 Length: 246  
 Score: 8.00 Matches: 8  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 1.68% Indels: 0  
 DB: 1 Gaps: 0

US-10-020-441-1 (1-1477) x MCTX\_MOUSE (1-246)

QY 1328 GGTGACAGTGTGACCTCTCCTC 1351

Db 200 GLYAspserGlyGlyProleuLeu 207

## RESULT 11

MCT1\_HUMAN STANDARD: PRT: 247 AA.

AC P23946; Q16018;

DT 01-MAR-1992 (Rel. 21, Created)

DT 01-MAR-1992 (Rel. 21, Last sequence update)

DT 15-JUN-2002 (Rel. 41, Last annotation update)

DE Chymase precursor (EC 3.4.21.39) (Mast cell protease I).

GN CM1 OR CTM OR CYH.

OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.

OX NCBI\_TaxID=9606;

RN [1]

RP SEQUENCE FROM N.A.

RX MEDLINE=91302311; PubMed=2071582;

RA Caughey G.H., Zerweck E.H., Vanderslice P.;

RA "Structure, chromosomal assignment, and deduced amino acid sequence

RT of a human gene for mast cell chymase."

RL J. Biol. Chem. 266:12956-12963(1991).

CC [2]

CC SEQUENCE FROM N.A.

RC TISSUE=Heart;

RX MEDLINE=91373329; PubMed=1894611;

RA Ureita H., Kinoshita A., Perez D.M., Misono K.S., Bumpus F.M.,

RA Graham R.M., Huseini A.,

RT "Cloning of the gene and cDNA for human heart chymase."

RL J. Biol. Chem. 266:17173-17179(1991).

CC [3]

CC SEQUENCE OF 22-247 FROM N.A.

RX MEDLINE=93265916; PubMed=8495723;

RA Suenaga Y., Kido H., Neki A., Enomoto M., Ishida K., Takagi K.,

RA Katunuma N.;

RT "Purification and molecular cloning of chymase from human tonsils."

RL FEBS Lett. 323:119-122(1993).

CC [4]

CC SEQUENCE OF 26-60 FROM N.A.

RC TISSUE=Placenta;

RX MEDLINE=91264818; PubMed=2049082;

RA "Gene D.E., Tschoep J.;

RT "Angiotensin II-forming heart chymase is a mast-cell-specific

RT enzyme."

RL Biochem. J. 276:567-568(1991).

CC [5]

CC X-RAY CRYSTALLOGRAPHY (1.9 ANGSTROMS).

RX MEDLINE=96062898; PubMed=9400368;

RA McGarrath M.E., Mirzadegan T., Schmidt B.F.;

RT "Crystal structure of phenylmethanesulfonyl fluoride-treated human

RT chymase at 1.9 A."

RL Biochemistry 36:14318-14324(1997).

CC [6]

CC X-RAY CRYSTALLOGRAPHY (2.2 ANGSTROMS).

RX MEDLINE=99134396; PubMed=9931257;

RA Pereira P.J.P., Wang Z.-M., Rubin H., Huber R., Bode W.,

RA Schechter N.M., Strobl S.;

RT "The 2.2-A crystal structure of human chymase in complex with

RT succinyl-Ala-Ala-Pro-Phe-chloromethylketone: structural explanation

RT for its dipeptidyl carboxypeptidase specificity."

RL J. Mol. Biol. 286:163-173(1999).

CC [7]





OY 1328 GGTGACAGTGTGACCTCTC 1351  
 |||||||||||||||||||  
 DB 201 GlyaspserglyProleuLeu 208

## RESULT 13

MCT1\_PAPHA STANDARD: PRT: 247 AA.  
 AC P52195:  
 DT 01-OCT-1996 (Rel. 34, Created)  
 DT 01-OCT-1996 (Rel. 34, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE Chymase precursor (EC 3.4.21.39) (Mast cell protease I).  
 GN CML OR CHM.  
 OS Papio hamadryas (Hamadryas baboon).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Cercopithecoidea;  
 OC Cercopithecinae; Papio.  
 OX NCBI\_TaxID=9557;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Liao Y., Karnik S., Husain A.;  
 RL Submitted (OCT-1995) to the EMBL/GenBank/DBJ databases.  
 CC -1- FUNCTION: MAJOR SECRETED PROTEASE OF MAST CELLS WITH SUSPECTED  
 CC ROLES IN VASOACTIVE PEPTIDE GENERATION, EXTRACELLULAR MATRIX  
 CC DEGRADATION, AND REGULATION OF GLAND SECRETION.  
 CC -1- CATALYTIC ACTIVITY: Preferential cleavage: Phe-I-Xaa > Tyr-I-Xaa >  
 CC Tfp-I-Xaa > Leu-I-Xaa.  
 CC -1- SUBCELLULAR LOCATION: MAST CELL GRANULES.  
 CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1. GRANZYME SUBFAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL: U38521; AAA91160.1; -;  
 DR EMBL: U38463; AAA91159.1; -;  
 DR HSSP: P23946; IKLT.  
 DR MEROPS: S01.140; -;  
 DR InterPro: IPR001314; Chymotrypsin.  
 DR InterPro: IPR001254; Ser\_protease\_Try.  
 DR Pfam: PF00089; trypsin.1  
 DR PRINTS: PR00722; CHYMOTRYPsin.  
 DR SMART: SM00020; TRYp-Spe; 1.  
 DR PROSITE: PS50240; TRYPsin\_DOM; 1.  
 DR PROSITE: PS00134; TRYPsin\_HIS; 1.  
 DR PROSITE: PS00135; TRYPsin\_SER; 1.  
 KW Hydrolyase; Serine protease; Glycoprotein; Zymogen; Signal.  
 FT SIGNAL 1 19  
 FT PROPEP 20 21  
 FT CHAIN 22 247  
 FT ACT\_SITE 66 66 CHARGE RELAY SYSTEM (BY SIMILARITY).  
 FT ACT\_SITE 110 110 CHARGE RELAY SYSTEM (BY SIMILARITY).  
 FT ACT\_SITE 203 203 CHARGE RELAY SYSTEM (BY SIMILARITY).  
 FT DISULFD 51 67  
 FT DISULFD 144 209  
 FT DISULFD 175 188 BY SIMILARITY.  
 FT DISULFD 175 188 BY SIMILARITY.  
 FT CARBOHD 80 80 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHD 103 103 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 SQ SEQUENCE 247 AA: 27339 MW: EDECLSEDFR72PDB8 CRC64;

## Alignment Scores:

Pred. No.: 9.15 Length: 247  
 Score: 8.00 Matches: 8  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 1.68% Indels: 0  
 DB: 1 Gaps: 0

US-10-020-441-1 (1-1477) x MCT1\_PAPHA (1-247)

OY 1328 GGTGACAGTGTGACCTCTC 1351  
 |||||||||||||||||||  
 DB 201 GlyaspserglyProleuLeu 208

## RESULT 14

MCT2\_MERUN STANDARD: PRT: 247 AA.  
 AC P50341:  
 DT 01-OCT-1996 (Rel. 34, Created)  
 DT 01-OCT-1996 (Rel. 34, Last sequence update)  
 DT 15-JUL-1998 (Rel. 36, Last annotation update)  
 DE Mast cell protease 2 precursor (EC 3.4.21.-).  
 OS Meriones unguiculatus (Mongolian jird) (Mongolian gerbil).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Gerbillinae;  
 OC Meriones.  
 OX NCBI\_TaxID=10047;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=MG/SEA; TISSUE=Intestine;  
 RA MEDLINE=96177868; PubMed=8615790;  
 RA Itoh H., Murakumo Y., Tomita M., Ide H., Kobayashi T., Maruyama H.,  
 RA Horii Y., Nawa Y.;  
 RT "Cloning of the cDNAs for mast-cell chymases from the jejunum of  
 RT Mongolian gerbils, Meriones unguiculatus, and their sequence  
 RT similarities with chymases expressed in the connective-tissue mast  
 RT cells of mice and rats."  
 RL Blochem. J. 314:923-929(1996).  
 CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1. GRANZYME SUBFAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL: D45174; BA08122.1; -;  
 DR HSSP: P23946; IKLT.  
 DR MEROPS: S01.140; -;  
 DR InterPro: IPR001314; Chymotrypsin.  
 DR InterPro: IPR001254; Ser\_protease\_Try.  
 DR Pfam: PF00089; trypsin.1  
 DR PRINTS: PR00722; CHYMOTRYPsin.  
 DR SMART: SM00020; TRYp-Spe; 1.  
 DR PROSITE: PS50240; TRYPsin\_DOM; 1.  
 DR PROSITE: PS00134; TRYPsin\_HIS; 1.  
 DR PROSITE: PS00135; TRYPsin\_SER; 1.  
 KW Hydrolyase; Serine protease; Signal; Glycoprotein; Zymogen;  
 KW Multigene family.  
 FT SIGNAL 1 19  
 FT PROPEP 20 21  
 FT CHAIN 22 247  
 FT ACT\_SITE 66 66 CHARGE RELAY SYSTEM (BY SIMILARITY).  
 FT ACT\_SITE 110 110 CHARGE RELAY SYSTEM (BY SIMILARITY).  
 FT ACT\_SITE 203 203 CHARGE RELAY SYSTEM (BY SIMILARITY).  
 FT DISULFD 51 67  
 FT DISULFD 144 209  
 FT DISULFD 175 188 BY SIMILARITY.  
 FT DISULFD 175 188 BY SIMILARITY.  
 FT CARBOHD 80 80 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHD 80 80 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 SQ SEQUENCE 247 AA: 27633 MW: CA737B5C043FBE30 CRC64;

## Alignment Scores:

Pred. No.: 9.15 Length: 247  
 Score: 8.00 Matches: 8  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 1.68% Indels: 0  
 DB: 1 Gaps: 0

US-10-020-441-1 (1-1477) x MCT2\_MERUN (1-247)

Qy 1328 GGTGACAGTGTGACCTCTCCTC 1351  
 |||||||  
 Db 201 GlyAspergilyglyProleuleu 208

RESULT 15  
 MCT2\_RAT  
 ID MCT2\_RAT STANDARD; PRT; 247 AA.  
 AC P00770;  
 DT 21-JUL-1986 (Rel. 01, Created)  
 DT 01-JAN-1988 (Rel. 06, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Mast cell protease II precursor (EC 3.4.21.-) (RMCP-II) (RMCP-2)  
 DE (Group-specific protease).  
 OS Rattus norvegicus (Rat).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
 OX NCBI\_TaxID=10116;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=87165980; PubMed=3549719;  
 RA Benfey P.N., Yin F.H., Leder P.;  
 RT "Cloning of the mast cell protease, RMCP II. Evidence for  
 RT cell-specific expression and a multi-gene family.";  
 RL J. Biol. Chem. 262:5377-5384(1987).  
 RN [2]  
 RP SEQUENCE OF 21-244.  
 RX MEDLINE=78124137; PubMed=629933;  
 RA Woodbury R.G., Katunuma N., Kobayashi K., Titani K., Neurath H.;  
 RT "Covalent structure of a group-specific protease from rat small  
 RT intestine. Appendix: crystallographic data for a group specific  
 RT protease from rat intestine.";  
 RL Biochemistry 17:811-819(1978).  
 RN [3]  
 RP X-RAY CRYSTALLOGRAPHY (1.9 ANGSTROMS).  
 RX MEDLINE=89166471; PubMed=323198;  
 RA Remington S.J., Woodbury R.G., Reynolds R.A., Matthews B.W.,  
 RA Neurath H.;  
 RT "The structure of rat mast cell protease II at 1.9-A resolution.";  
 RL Biochemistry 27:8097-8105(1988).  
 CC -1- FUNCTION: THIS ENZYME, ISOLATED FROM SMALL INTESTINE, SPECIFICALLY  
 CC INACTIVATES THE APO FORMS OF A CERTAIN GROUP OF INTRACELLULAR  
 CC PYRIDOXAL PHOSPHATE-REQUIRING ENZYMES. IT HAS CHYMOTRYPSIN-LIKE  
 CC SPECIFICITY TOWARDS SMALL SUBSTRATES.  
 CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1. GRANZYME SUBFAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL: J02712; AAA66284.1; -  
 CC DR PIR: A00957; PRRG.  
 DR PIR: A29548; A29548.  
 DR PDB: 3RP2; 09-JAN-89.  
 DR MEROPS: S01.141; -  
 DR InterPro: IPR001314; Chymotrypsin.  
 DR InterPro: IPR001254; Ser.protease\_Try.  
 DR Pfam: PF00089; trypsin\_1  
 DR PRINTS: PR00722; CHYMOTRYPSIN.  
 DR SMART: SM00020; Tryp\_Spc; 1.  
 DR PROSITE: PS50240; TRYPSIN\_DOM; 1.  
 DR PROSITE: PS00134; TRYPSIN\_HIS; 1.  
 DR PROSITE: PS00135; TRYPSIN\_SER; 1.  
 KW Hydrolyase; Serine protease; 3d-structure; Zymogen; Signal;  
 KM Multigene family.  
 FT SIGNAL 1 18  
 FT PROPEP 19 20 ACTIVATION PEPTIDE.  
 FT CHAIN 21 247 MAST CELL PROTEASE II.  
 FT ACT\_SITE 65 65 CHARGE RELAY SYSTEM.

FT ACT\_SITE 109 109  
 FT ACT\_SITE 202 202 CHARGE RELAY SYSTEM.  
 FT DISULFID 50 66 CHARGE RELAY SYSTEM.  
 FT DISULFID 143 208  
 FT DISULFID 174 187  
 FT DISULFID 174 187  
 FT CONFLICT 238 238  
 FT STRAND 22 22 W -> T (IN REF. 2).  
 FT STRAND 22 22  
 FT STRAND 25 26  
 FT STRAND 29 30  
 FT TURN 29 30  
 FT TURN 33 34  
 FT STRAND 35 41  
 FT TURN 43 44  
 FT STRAND 47 54  
 FT STRAND 59 62  
 FT STRAND 64 66  
 FT HELIX 70 75  
 FT STRAND 79 79  
 FT STRAND 80 81  
 FT TURN 85 86  
 FT STRAND 88 97  
 FT TURN 99 100  
 FT STRAND 111 115  
 FT STRAND 122 122  
 FT TURN 123 124  
 FT STRAND 125 125  
 FT TURN 133 134  
 FT TURN 139 140  
 FT STRAND 142 147  
 FT STRAND 150 152  
 FT TURN 153 154  
 FT STRAND 155 157  
 FT STRAND 160 160  
 FT STRAND 162 169  
 FT HELIX 171 173  
 FT TURN 174 178  
 FT TURN 182 184  
 FT STRAND 185 188  
 FT TURN 191 192  
 FT STRAND 196 196  
 FT TURN 199 203  
 FT STRAND 205 208  
 FT TURN 209 210  
 FT STRAND 211 218  
 FT TURN 221 222  
 FT STRAND 227 231  
 FT HELIX 232 243  
 SQ SEQUENCE 247 AA; 27101 MW; 051988042A97A47E CRC64;

Alignment Scores:  
 Pred. No.: 9.15  
 Score: 8.00  
 Percent Similarity: 100.00%  
 Best Local Similarity: 100.00%  
 Query Match: 1.68%  
 DB: 1 Gaps: 0

US-10-020-441-1 (1-1477) x MCT2\_RAT (1-247)  
 Qy 1328 GGTGACAGTGTGACCTCTCCTC 1351  
 |||||||  
 Db 200 GlyAspergilyglyProleuleu 207

RESULT 16  
 MCT3\_RAT  
 ID MCT3\_RAT STANDARD; PRT; 247 AA.  
 AC P50339;  
 DT 01-OCT-1996 (Rel. 34, Created)  
 DT 01-OCT-1996 (Rel. 34, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Mast cell protease III precursor (EC 3.4.21.-) (RMCP-III) (RMCP-3).  
 OS Rattus norvegicus (Rat).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.

```

OX NCBI_TaxID-10116;
RN [1]
RN SEQUENCE FROM N.A.
RC TISSUE-Pertoneal mast cells;
RX MEDLINE-96033070; PubMed-7487912;
RA Ide H., Itoh H., Tomita M., Murakumo Y., Kobayashi T.,
RA Maryama H., Osada Y., Nawa Y.;
RT "Cloning of the cDNA encoding a novel rat mast-cell proteinase,
RT RMC-3, and its expression in comparison with other rat mast-cell
RT proteinases."
RL Biochem. J. 311:675-680(1995).
CC -1- SUBCELLULAR LOCATION: Secretory granules.
CC -1- TISSUE SPECIFICITY: MAST CELLS.
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1. GRANZYME SUBFAMILY.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
CC
DR EMBL: D38495; BAA07507.1; -.
DR HSSP: P23946; 1KLT.
DR MEROPS: S01.150; -.
DR InterPro: IPR001314; Chymotrypsin.
DR InterPro: IPR001254; Ser_protease_Try.
DR Pfam: PF00089; trypsin; 1.
DR PRINTS: PR00722; CHYMOTRYPSIN.
DR SMART: SM0020; TRYP-SPC; 1.
DR PROSITE: PS50240; TRYPSIN_DOM; 1.
DR PROSITE: PS00134; TRYPSIN_HIS; 1.
DR PROSITE: PS00135; TRYPSIN_SER; 1.
KM Hydroxylase; Serine protease; Signal; Glycoprotein; Zymogen;
KM Multigene family.
FT SIGNAL 1 19 POTENTIAL.
FT PROPEP 20 21 ACTIVATION PEPTIDE.
FT CHAIN 22 247 MAST CELL PROTEASE III.
FT ACT_SITE 66 66 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE 110 110 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE 203 203 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT DISULFID 51 67 BY SIMILARITY.
FT DISULFID 144 209 BY SIMILARITY.
FT DISULFID 175 188 BY SIMILARITY.
FT CARBOHD 80 80 N-LINKED (GLCNAC... ) (POTENTIAL).
SQ SEQUENCE 247 AA; 27569 MW; 6525DBF1BFD053 CRC64;

Alignment Scores:
Pred. No.: 9.15 Length: 247
Score: 8.00 Matches: 8
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 1.68% Indels: 0
DB: 1 Gaps: 0

US-10-020-441-1 (1-1477) x MCT3_RAT (1-247)
QY 1328 GGTCACAGTGTCGACCTCTCTC 1351
DB 201 GLYASPSERGlyGlyProleuLeu 208

RESULT 17
MCT5_MOUSE
ID MCT5_MOUSE STANDARD; PRT; 247 AA.
AC P21844;
DR 01-MAY-1991 (Rel. 18, Created)
DR 01-DEC-1992 (Rel. 24, Last sequence update)
DR 15-JUN-2002 (Rel. 41, Last annotation update)
DE Mast cell protease 5 precursor (EC 3.4.21.-) (MMCP-5) (Mast cell
DE chymase 1).
GN MCP5.
OS Mus musculus (Mouse).

```

```

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID-10090;
RN [1]
RN SEQUENCE FROM N.A.
RX MEDLINE-92041862; PubMed-1939089;
RA McNeil H.P., Austen K.F., Somerville L.L., Gurish M.F., Stevens R.L.;
RT "Molecular cloning of the mouse mast cell protease-5 gene. A novel
RT secretory granule protease expressed early in the differentiation of
RT serosal mast cells."
RL J. Biol. Chem. 266:20316-20322(1991).
RN [2]
RN SEQUENCE FROM N.A.
RC STRAIN-leadin X AL; TISSUE-connective tissue.
RX MEDLINE-91285010; PubMed-2060576;
RA Huang R., Blom T., Hellman L.;
RT "Cloning and structural analysis of MMCP-1, MMCP-4 and MMCP-5, three
RT mouse mast cell-specific serine proteases."
RL Eur. J. Immunol. 21:1611-1621(1991).
RN [3]
RN SEQUENCE FROM N.A.
RX MEDLINE-92287966; PubMed-1376147;
RA Chu W., Johnson D.A., Musich P.R.;
RT "Molecular cloning and characterization of mouse mast cell chymases."
RL Biochim. Biophys. Acta 1121:83-87(1992).
RN [4]
RN SEQUENCE OF 22-51.
RX MEDLINE-90222202; PubMed-2326280;
RA Reynolds D.S., Stevens R.L., Lane W.S., Carr M.H., Austen K.F.,
RA Serafini W.E.;
RT "Different mouse mast cell populations express various combinations
RT of at least six distinct mast cell serine proteases."
RL Proc. Natl. Acad. Sci. U.S.A. 87:3230-3234(1990).
RN [5]
RN SUBCELLULAR LOCATION.
RX MEDLINE-92407343; PubMed-1527387;
RA McNeil H.P., Frenkel D.P., Austen F., Friend D.S., Stevens R.L.;
RT "Translation and granule localization of mouse mast cell protease-5.
RT Immunodetection with specific antipeptide Ig."
RL J. Immunol. 149:2466-2472(1992).
CC -1- SUBCELLULAR LOCATION: Secretory granules.
CC -1- TISSUE SPECIFICITY: MAST CELLS.
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1. GRANZYME SUBFAMILY.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
CC
DR EMBL: X68805; CAA48705.1; ALT_INIT.
DR EMBL: M73759; AAA40105.1; -.
DR EMBL: M73760; -. NOT_ANNOTATED_CDs.
DR EMBL: M68898; AAA39492.1; -.
DR PIR: A41076; A41076.
DR PIR: C35646; C35646.
DR HSSP: P23946; 1KLT.
DR MEROPS: S01.150; -.
DR MGD: MGI:96941; MCP5.
DR InterPro: IPR001314; Chymotrypsin.
DR InterPro: IPR001254; Ser_protease_Try.
DR Pfam: PF00089; trypsin; 1.
DR PRINTS: PR00722; CHYMOTRYPSIN.
DR SMART: SM0020; TRYP-SPC; 1.
DR PROSITE: PS50240; TRYPSIN_DOM; 1.
DR PROSITE: PS00134; TRYPSIN_HIS; 1.
DR PROSITE: PS00135; TRYPSIN_SER; 1.
KM Hydroxylase; Serine protease; Signal; Glycoprotein; Zymogen.
FT SIGNAL 1 19 POTENTIAL.
FT PROPEP 20 21 ACTIVATION PEPTIDE.
FT CHAIN 22 247 MAST CELL PROTEASE 5.

```

```
FT ACT_SITE 66 66 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE 110 110 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE 203 203 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT DISULFID 51 67 BY SIMILARITY.
FT DISULFID 144 209 BY SIMILARITY.
FT DISULFID 175 188 BY SIMILARITY.
FT CARBOHYD 80 80 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CONFLICT 5 5 T -> A (IN REF. 1).
FT CONFLICT 51 51 C -> R (IN REF. 4).
FT CONFLICT 224 224 A -> R (IN REF. 3).
SQ SEQUENCE 247 AA; 27586 MW; 24C290CF61237DC7 CRC64;

Alignment Scores:
Pred. No.: 9.15 Length: 247
Score: 8.00 Matches: 8
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 1.68% Indels: 0
Gaps: 0

US-10-020-441-1 (1-1477) x MCT5_MOUSE (1-247)
QY 1328 GGTGACAGTGTGACCTCTCTC 1351
Db 201 GlyAspserGlyGlyProLeuLeu 208

RESULT 18
MCT1_CANFA STANDARD: PRT: 249 AA.
ID MCT1_CANFA
AC P21842;
DT 01-MAY-1991 (Rel. 18, Created)
DT 01-MAY-1991 (Rel. 18, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Chymase precursor (EC 3.4.21.39) (Mast cell protease I).
OS Canis familiaris (Dog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.
OX NCBI_TaxID=9615;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=90335214; PubMed=2378872;
RA Caughey G.H., Raymond W.W., Vanderslice P.;
RT "Dog mast cell chymase: molecular cloning and characterization.";
RL Biochemistry 29:5166-5171(1990).
RN [2]
RP SEQUENCE FROM N.A.
RX STRAIN=Beagle;
RA MEDLINE=98026156; PubMed=9379034;
RA Caughey G.H., Blount J.L., Koerber K.L., Kitamura M., Fang K.C.;
RT "Cloning and expression of the dog mast cell alpha-chymase gene.";
RL J. Immunol. 159:4367-4375(1997).
RN [3]
RP SEQUENCE OF 22-46.
RX MEDLINE=88107816; PubMed=3122835;
RA Caughey G.H., Viro N.F., Lazarus S.C., Nagel J.A.;
RT "Purification and characterization of dog mastocytoma chymase:
RT identification of an octapeptide conserved in chymotryptic leukocyte
RT proteinases.";
RL Biochim. Biophys. Acta 952:142-149(1988).
RN [4]
RP SEQUENCE FROM N.A.
RX MEDLINE=952142149(1988).
RA "CATALYTIC ACTIVITY: Preferential cleavage: Phe-I-Xaa > Tyr-I-Xaa >
RA Trp-I-Xaa > Leu-I-Xaa.
RN [5]
RP -I- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1. GRANZYME SUBFAMILY.
RN [6]
RP This SWISS-PROT entry is copyright. It is produced through a collaboration
RN between the Swiss Institute of Bioinformatics and the EMBL outstation -
RN the European Bioinformatics Institute. There are no restrictions on its
RN use by non-profit institutions as long as its content is in no way
RN modified and this statement is not removed. Usage by and for commercial
RN entities requires a license agreement (See http://www.isb-sib.ch/announce/
RN or send an email to license@isb-sib.ch).
RN [7]
RP EMBL; J02904; AAA30835.1; -
RN EMBL; U89607; AAB94641.1; -
```

```
DR PIR: A35842; A35842.
DR HSSP: P23946; IKLT.
DR MEROPS: S01.140; -.
DR InterPro: IPR001314; Chymotrypsin.
DR InterPro: IPR001254; Ser.protease_Try.
DR Pfam: PF00089; trypsin. 1.
DR PRINTS: PR00722; CHYMOTRYPSIN.
DR SMART: SM00020; TRYP-Ser. 1.
DR PROSITE: PS50240; TRYP-SIN_DOM. 1.
DR PROSITE: PS00134; TRYP-SIN_HIS. 1.
DR PROSITE: PS00135; TRYP-SIN_SER. 1.
KW Hydrolase; Serine protease; zymogen; Signal.
FT SIGNAL 1 19
FT PROPEP 20 21 ACTIVATION PEPTIDE.
FT CHAIN 22 249 CHYMASE.
FT ACT_SITE 66 66 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE 110 110 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE 203 203 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT DISULFID 51 67 BY SIMILARITY.
FT DISULFID 144 209 BY SIMILARITY.
FT DISULFID 175 188 BY SIMILARITY.
FT CONFLICT 29 29 K -> R (IN REF. 3).
FT CONFLICT 38 38 H -> Q (IN REF. 3).
FT CONFLICT 45 45 R -> T (IN REF. 3).
SQ SEQUENCE 249 AA; 27811 MW; 3BBD0A6C2855F540 CRC64;

Alignment Scores:
Pred. No.: 9.14 Length: 249
Score: 8.00 Matches: 8
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 1.68% Indels: 0
Gaps: 0

US-10-020-441-1 (1-1477) x MCT1_CANFA (1-249)
QY 1328 GGTGACAGTGTGACCTCTCTC 1351
Db 201 GlyAspserGlyGlyProLeuLeu 208

RESULT 19
CATG_HUMAN STANDARD: PRT: 255 AA.
ID CATG_HUMAN
AC P08311;
DT 01-AUG-1988 (Rel. 08, Created)
DT 01-JAN-1990 (Rel. 13, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Cathepsin G precursor (EC 3.4.21.20) (CG).
GN CTSG.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=89340411; PubMed=2569462;
RA Hohn P.A., Popescu N.C., Hanson R.D., Salvesen G., Ley T.J.;
RT "Genomic organization and chromosomal localization of the human
RT cathepsin G gene.";
RL J. Biol. Chem. 264:13412-13419(1989).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=87299663; PubMed=3304423;
RA Salvesen G., Farley D., Shuman J., Przybyla A., Reilly C., Travis J.;
RT "Molecular cloning of human cathepsin G: structural similarity to
RT mast cell and cytotoxic T lymphocyte proteinases.";
RL Biochemistry 26:2289-2293(1987).
RN [3]
RP SEQUENCE FROM N.A.
RX STRAUBERG R.;
RN submitted (SEP-2001) to the EMBL/GenBank/DBJ databases.
RN [4]
```

RP SEQUENCE OF 21-45.  
 RX MEDLINE=87097924; PubMed=3799965;  
 RA Heck L.W., Rostand K.S., Hunter F.A., Brown A.;  
 RT "Isolation, characterization, and amino-terminal amino acid sequence  
 RT analysis of human neutrophil cathepsin G from normal donors.";  
 RL Anal. Biochem. 158:217-227(1986).  
 RN [5]  
 RP SEQUENCE OF 21-36.  
 RX MEDLINE=89315847; PubMed=2501794;  
 RA Gabay J.E., Scott R.W., Campanelli D., Griffith J., Wilde C.,  
 RA "Marra M.N., Seeger M., Nathan C.F.;  
 RT "Antibiotic proteins of human polymorphonuclear leukocytes.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 86:5610-5614(1989).  
 RN [6]  
 RP X-RAY CRYSTALLOGRAPHY (1.8 ANGSTROMS).  
 RX MEDLINE=97051807; PubMed=8896442;  
 RA Hof P., Mayr I., Huber R., Korzus E., Potempa J., Travis J.,  
 RA "The 1.8 A crystal structure of human cathepsin G in complex with  
 RT Suc-Val-Pro-Phep-(Oph)<sub>2</sub>: a Janus-faced proteinase with two opposite  
 RT specificities.";  
 RL EMBO J. 15:5481-5491(1996).  
 RN [7]  
 RP X-RAY CRYSTALLOGRAPHY (1.9 ANGSTROMS).  
 RA Medrano F.J., Bode W., Banbula A., Potempa J.;  
 RL Submitted (SEP-1997) to the PDB data bank.  
 RN [8]  
 RP VARIANT SER-125.  
 RX MEDLINE=93202661; PubMed=8454293;  
 RA Lueddecke B., Poller W., Olek K., Bartholome K.;  
 RT "Sequence variant of the human cathepsin G gene.";  
 RL Hum. Genet. 91:83-84(1993).  
 CC -1- CATALYTIC ACTIVITY: SPECIFICITY SIMILAR TO CHYMOTRYPSIN.  
 CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL: M16117; AAA52126.1; -;  
 DR EMBL: J04990; AAA51919.1; -;  
 DR EMBL: BC014460; AAH14460.1; -;  
 DR PIR: A32627; A32627.  
 DR PIR: A27122; A27122.  
 DR PIR: A05307; A05307.  
 DR PDB: 1CGH; 07-JUL-97.  
 DR PDB: 1A08; 14-OCT-98.  
 DR MEROPS: S01.133; -;  
 DR GeneW: HGNC:2532; CTSG.  
 DR MIM: 116830; -;  
 DR InterPro: IPR001314; Chymotrypsin.  
 DR InterPro: IPR001254; Ser.protease\_Try.  
 DR Pfam: PF00089; trypsin.1.  
 DR PRINTS: PR00722; CHYMOTRYPSIN.  
 DR SMART: SM00020; TRY-Spec.1.  
 DR PROSITE: PS50240; TRYPSIN\_DOM; 1.  
 DR PROSITE: PS00134; TRYPSIN\_HIS; 1.  
 DR PROSITE: PS00135; TRYPSIN\_SER; 1.  
 KW Hydrolase; Serine protease; Zymogen; Glycoprotein; Signal;  
 KW Polymorphism; 3D-structure.  
 FT SIGNAL 1 18  
 FT PROPEP 19 20 ACTIVATION PEPTIDE.  
 FT CHAIN 21 255 CATHEPSIN G.  
 FT ACT\_SITE 64 64 CHARGE RELAY SISTEM.  
 FT ACT\_SITE 108 108 CHARGE RELAY SISTEM.  
 FT ACT\_SITE 201 201 CHARGE RELAY SISTEM.  
 FT DISULFID 49 65  
 FT DISULFID 142 207  
 FT DISULFID 172 186

FT CARBOHYD 71 71 N-LINKED (GLCNAC. . .).  
 FT VARIANT 125 125 N -> S.  
 FT CONFLICT 39 39 /FTID=VAR\_006491.  
 FT CONFLICT 41 43 Q -> E (IN REF. 2).  
 FT CONFLICT 41 43 OSP -> TSG (IN REF. 2).  
 SQ SEQUENCE 255 AA; 28837 MW; 6228E741E6A43889 CRC64;  
 Alignment Scores:  
 Pred. No.: 9.1 Length: 255  
 Score: 8.00 Matches: 8  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 1.68% Indels: 0  
 DB: 1 Gaps: 0  
 US-10-020-441-1 (1-1477) x CATG\_HUMAN (1-255)  
 QY 1328 GGTGACAGTGTGACCTCTC 1351  
 DB 199 GLYAspsercLYGlyProleu 206  
 RESULT 20  
 KLC\_RAT  
 ID KLC\_RAT STANDARD: PRT; 259 AA.  
 AC P36376;  
 DT 01-JUN-1994 (Rel. 29. Created)  
 DT 01-JUN-1994 (Rel. 29, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Glandular kallikrein 12, submandibular/renal precursor (EC 3.4.21.35)  
 DE (Tissue kallikrein) (RSKG-3).  
 GN KLK12 OR KLK-12.  
 OS Rattus norvegicus (Rat).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
 OC NCBI\_TaxID=10116;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=8908074; PubMed=2849988;  
 RA Chen Y.-P., Chao J., Chao L.;  
 RT "Molecular cloning and characterization of two rat renal kallikrein  
 RT genes";  
 RL Biochemistry 27:7189-7196(1988).  
 CC -1- FUNCTION: GLANDULAR KALLIKREINS CLEAVE MET-LYS AND ARG-SER BONDS  
 CC IN KININOGEN TO RELEASE LYS-BRADYKININ.  
 CC -1- CATALYTIC ACTIVITY: Preferential cleavage of Arg-I-Xaa bonds in  
 CC small molecule substrates. Highly selective action to release  
 CC kallidin (lysyl-bradykinin) from kininogen involves hydrolysis of  
 CC Met-I-Xaa or Leu-I-Xaa.  
 CC -1- TISSUE SPECIFICITY: KIDNEY AND SUBMANDIBULAR GLAND.  
 CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1. KALLIKREIN SUBFAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL: M19648; AAA51640.1; -;  
 DR EMBL: M22922; AAA51640.1; JOINED.  
 DR PIR: B31136; B31136.  
 DR HSP: P00759; ITON.  
 DR MEROPS: S01.160; -;  
 DR InterPro: IPR001314; Chymotrypsin.  
 DR InterPro: IPR001254; Ser.protease\_Try.  
 DR Pfam: PF00089; trypsin.1.  
 DR PRINTS: PR00722; CHYMOTRYPSIN.  
 DR SMART: SM00020; TRY-Spec.1.  
 DR PROSITE: PS50240; TRYPSIN\_DOM; 1.  
 DR PROSITE: PS00134; TRYPSIN\_HIS; 1.  
 DR PROSITE: PS00135; TRYPSIN\_SER; 1.  
 KW Hydrolase; Serine protease; Glycoprotein; Multigene family; Zymogen;

KW Signal. 1 18 PROBABLE.  
FT PROPEP 19 24 ACTIVATION PEPTIDE (PROBABLE).  
FT CHAIN 25 259 GLANDULAR KALLIKREIN 12,  
SUBMANDIBULAR/RENAL.  
FT ACT\_SITE 63 63 CHARGE RELAY SYSTEM.  
FT ACT\_SITE 118 118 CHARGE RELAY SYSTEM.  
FT ACT\_SITE 211 211 CHARGE RELAY SYSTEM.  
FT DISULFID 31 171 BY SIMILARITY.  
FT DISULFID 48 64 BY SIMILARITY.  
FT DISULFID 150 217 BY SIMILARITY.  
FT DISULFID 182 196 BY SIMILARITY.  
FT DISULFID 207 232 BY SIMILARITY.  
FT CARBOHYD 91 91 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 106 106 N-LINKED (GLCNAC. . .) (POTENTIAL).  
SQ SEQUENCE 259 AA; 28759 MW; 535A8E25435144F CRC64;

Alignment Scores:  
Pred. No.: 9.08 Length: 259  
Score: 8.00 Matches: 8  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 1.68% Indels: 0  
DB: 1 Gaps: 0

US-10-020-441-1 (1-1477) x KLK7\_RAT (1-259)  
OY 1328 GGTGACAGTGTGACCTCTCCTC 1351  
|||||  
Db 209 GlyasperglyglyProLeuLeu 216

RESULT 21  
KLK7\_RAT STANDARD; PRT; 261 AA.  
AC P36373;  
DT 01-JUN-1994 (Rel. 29, Created)  
DT 01-JUN-1994 (Rel. 29, Last sequence update)  
DT 15-JUN-1999 (Rel. 38, Last annotation update)  
DE Glandular kallikrein 7, submandibular/renal precursor (EC 3.4.21.35)  
DE (tissue kallikrein) (R0K-7) (esterase B) (Proteinase A).  
GN KLK7 OR KLK-7.  
OS Rattus norvegicus (Rat).  
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Rattus.  
OX NCBI\_TaxID=10116;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=89088074; PubMed=2849988;  
RA Chen Y.-P., Chao J., Chao L.;  
RT "Molecular cloning and characterization of two rat renal kallikrein  
genes.";  
RL Biochemistry 27:7189-7196(1988).  
RN [2]  
RP SEQUENCE OF 25-75.  
RC TISSUE-Submaxillary gland;  
RX MEDLINE=88198057; PubMed=3482210;  
RA Kato H., Nakamishi E., Enjyoji K., Hayashi I., Oh-Ishi S., Iwanaga S.;  
RT "Characterization of serine proteinases isolated from rat  
submaxillary gland: with special reference to the degradation of rat  
kininogens by these enzymes.";  
RT J. Biochem. 102:1389-1404(1987).  
RL [1]  
CC -1- FUNCTION: GLANDULAR KALLIKREINS CLEAVE MET-LYS AND ARG-SER BONDS  
IN KININOGEN TO RELEASE LYS-BRADYKININ.  
CC -1- CATALYTIC ACTIVITY: Preferential cleavage of Arg-I-Xaa bonds in  
small molecule substrates. Highly selective action to release  
kallidin (lysyl-bradykinin) from kininogen involves hydrolysis of  
Met-I-Xaa or Leu-I-Xaa.  
CC -1- TISSUE SPECIFICITY: KIDNEY AND SUBMANDIBULAR GLAND.  
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1. KALLIKREIN SUBFAMILY.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
CC -----  
DR EMBL: M19647; AAAA1461.1; -.  
DR PIR: A31136; A31136.  
DR PIR: B41429; B41429.  
DR HSSP: P00759; ITON.  
DR MEROPS: S01.406; -.  
DR InterPro: IPR001314; Chymotrypsin.  
DR InterPro: IPR001254; Ser\_protease\_Try.  
DR Pfam: PF00089; Trypsin.1.  
DR PRINTS: PR00722; CHYMOTRYPSIN.  
DR SMART: SM00020; TRYD\_SPE; 1.  
DR PROSITE: PS02040; TRYPsin\_DOM; 1.  
DR PROSITE: PS00134; TRYPsin\_HIS; 1.  
DR PROSITE: PS00135; TRYPsin\_SER; 1.  
KW Hydrolyase; Serine protease; Glycoprotein; Multigene family; Zymogen;  
KW Signal.  
FT SIGNAL. 1 18 PROBABLE.  
FT PROPEP 19 24 ACTIVATION PEPTIDE (PROBABLE).  
FT CHAIN 25 261 GLANDULAR KALLIKREIN 7,  
SUBMANDIBULAR/RENAL.  
FT ACT\_SITE 65 65 CHARGE RELAY SYSTEM.  
FT ACT\_SITE 120 120 CHARGE RELAY SYSTEM.  
FT ACT\_SITE 213 213 CHARGE RELAY SYSTEM.  
FT DISULFID 31 173 BY SIMILARITY.  
FT DISULFID 50 66 BY SIMILARITY.  
FT DISULFID 152 219 BY SIMILARITY.  
FT DISULFID 184 198 BY SIMILARITY.  
FT DISULFID 209 234 BY SIMILARITY.  
FT DISULFID 209 234 BY SIMILARITY.  
FT CARBOHYD 108 108 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 108 108 S-> D (IN REF. 2).  
FT CONFLICT 35 35 T-> S (IN REF. 2).  
FT CONFLICT 46 46  
SQ SEQUENCE 261 AA; 28972 MW; 4FB06C422EF25AF16 CRC64;

Alignment Scores:  
Pred. No.: 9.07 Length: 261  
Score: 8.00 Matches: 8  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 1.68% Indels: 0  
DB: 1 Gaps: 0

US-10-020-441-1 (1-1477) x KLK7\_RAT (1-261)  
OY 1328 GGTGACAGTGTGACCTCTCCTC 1351  
|||||  
Db 211 GlyasperglyglyProLeuLeu 218

RESULT 22  
PRTB\_STRGR  
ID PRTB\_STRGR STANDARD; PRT; 299 AA.  
AC P00777;  
DT 21-JUL-1986 (Rel. 01, Created)  
DT 01-FEB-1991 (Rel. 17, Last sequence update)  
DT 15-JUN-2002 (Rel. 41, Last annotation update)  
DE Streptomyces griseus B precursor (EC 3.4.21.81) (Protease B) (SGPB) (Pronase  
enzyme B).  
GN SGPB.  
OS Streptomyces griseus.  
OC Bacteria; Actinobacteria; Actinobacteria (class); Actinobacteridae;  
OC Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomycetes.  
OX NCBI\_TaxID=1911;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=87279934; PubMed=3112129;  
RA Henderson G., Krzygman P., Liu C.J., Davey C.C., Malek L.T.;  
RT "Characterization and structure of genes for proteases A and B from  
Streptomyces griseus.";  
RL J. Bacteriol. 169:3778-3784(1987).  
RN [2]

RP SEQUENCE OF 115-299.  
 RC STRAIN-K-1;  
 RX MEDLINE-75127477; PubMed-4218101;  
 RA Jurassek L., Carpenter M.R., Smillie L.B., Gertler A., Levy S.,  
 RT Ericsson L.H.;  
 RT "Amino acid sequence of Streptomyces griseus protease B, A MAJOR  
 COMPONENT OF PRONASE."  
 RL Biochem. Biophys. Res. Commun. 61:1095-1100(1974).  
 RN [3]  
 RP X-RAY CRYSTALLOGRAPHY (2.8 ANGSTROMS), AND REVISIONS.  
 RX MEDLINE-76051298; PubMed-1186854;  
 RA Delbaere L.T.J., Hutcheon W.L.B., James M.N.G., Thiesen W.E.;  
 RT "Tertiary structural differences between microbial serine proteases  
 and pancreatic serine enzymes."  
 RL Nature 257:758-763(1975).  
 RN [4]  
 RP X-RAY CRYSTALLOGRAPHY (2.8 ANGSTROMS), AND REVISIONS.  
 RX MEDLINE-79210651; PubMed-110426;  
 RA Delbaere L.T.J., Brayer G.D., James M.N.G.;  
 RT "The 2.8-A resolution structure of Streptomyces griseus protease B  
 and its homology with alpha-chymotrypsin and Streptomyces griseus  
 protease A."  
 RL Can. J. Biochem. 57:135-144(1979).  
 RN [5]  
 RP X-RAY CRYSTALLOGRAPHY (2.1 ANGSTROMS).  
 RX MEDLINE-89178636; PubMed-2494344;  
 RA Greenblatt H.M., Ryan C.A., James M.N.G.;  
 RT "Structure of the complex of Streptomyces griseus proteinase B and  
 polypeptide chymotrypsin inhibitor-1 from Russet Burbank potato  
 tubers at 2.1-A resolution."  
 RL J. Mol. Biol. 205:201-228(1989).  
 RN [6]  
 RP ACTIVE SITE.  
 RX MEDLINE-74302902; PubMed-4212092;  
 RA Gertler A.;  
 RT "Inhibition of Streptomyces griseus protease B by peptide  
 chloromethyl ketones: partial mapping of the binding site and  
 identification of the reactive residue."  
 RL FEBS Lett. 43:81-85(1974).  
 RN [7]  
 RP ACTIVE SITE.  
 RX MEDLINE-68124983; PubMed-5636372;  
 RA Wahlby S., Engstrom L.;  
 RT "Studies on Streptomyces griseus protease. II. The amino acid  
 sequence around the reactive serine residue of DFP-sensitive  
 components with esterase activity."  
 RL Biochim. Biophys. Acta 151:402-408(1968).  
 CC -1- FUNCTION: HAS A PRIMARY SPECIFICITY FOR LARGE ALIPHATIC OR  
 AROMATIC AMINO ACIDS.  
 CC -1- CATALYTIC ACTIVITY: Hydrolysis of proteins with specificity  
 similar to trypsin.  
 CC -1- SUBUNIT: MONOMER.  
 CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S2A.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL: M17104; AAA26819.1; -;  
 DR EMBL: A24973; CAA01747.1; -;  
 DR PIR: A00964; PRSMBG.  
 DR PIR: B26974; B26974.  
 DR PDB: 3SG8; 15-JUL-90.  
 DR PDB: 4SG8; 15-JUL-90.  
 DR PDB: 1SGP; 15-OCT-95.  
 DR PDB: 1SGO; 15-OCT-95.  
 DR PDB: 1SGR; 15-OCT-95.  
 DR MEROPS: S01.262; -;  
 DR InterPro: IPR004236; AL-protease.

DR InterPro: IPR001316; Endopdpses2A.  
 DR InterPro: IPR001254; Ser.protease\_Try.  
 DR Pfam: PF00089; trypsin\_1.  
 DR Pfam: PF02983; AL-protease\_1.  
 DR PRINTS: PR00861; ALTYICTPASE.  
 DR SMART: SM00020; TRY-SPC; 1.  
 DR PROSITE: PS00134; TRYPsin\_HIS; 1.  
 DR PROSITE: PS00135; TRYPsin\_SER; 1.  
 KW Hydrolyase; Serine protease; Zymogen; Signal; 3D-structure.  
 FT SIGNAL 1 38  
 FT PROPEP 39 114  
 FT CHAIN 115 299 STREPTOGRISIN B.  
 FT DISULFID 128 148  
 FT DISULFID 249 276  
 FT ACT\_SITE 147 147  
 FT ACT\_SITE 177 177  
 FT ACT\_SITE 255 255 CHARGE RELAY SYSTEM.  
 FT CONFLICT 153 153 T -> TG (IN REF. 2).  
 FT CONFLICT 157 157 MISSING (IN REF. 2).  
 FT CONFLICT 293 293 A -> V (IN REF. 2).  
 FT CONFLICT 296 296 V -> A (IN REF. 2).  
 FT STRAND 117 117  
 FT TURN 118 118  
 FT STRAND 120 123  
 FT TURN 124 125  
 FT STRAND 126 129  
 FT STRAND 132 136  
 FT TURN 137 138  
 FT STRAND 139 144  
 FT HELIX 146 149  
 FT TURN 150 151  
 FT STRAND 154 156  
 FT TURN 159 160  
 FT STRAND 164 172  
 FT STRAND 175 175  
 FT STRAND 178 183  
 FT STRAND 193 194  
 FT TURN 195 196  
 FT STRAND 197 198  
 FT STRAND 201 202  
 FT TURN 207 208  
 FT STRAND 210 215  
 FT TURN 216 218  
 FT STRAND 219 235  
 FT TURN 236 237  
 FT STRAND 238 246  
 FT TURN 252 253  
 FT TURN 255 256  
 FT TURN 258 261  
 FT TURN 262 263  
 FT STRAND 264 275  
 FT TURN 276 278  
 FT STRAND 279 286  
 FT HELIX 287 294  
 FT TURN 295 295  
 FT STRAND 297 298  
 SQ SEQUENCE 299 AA: 30554 MW: 7A272C9516C7B9B1 CRC64: ?  
 Alignment Scores:  
 Pred. No.: 8.87 Length: 299  
 Score: 8.00 Matches: 8  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 1.68% Indels: 0  
 DB: 1 Gaps: 0  
 US-10-020-441-1 (1-1477) x PRTB\_STRGR (1-299)  
 QY 1325 CCAGGTGACAGTGTGACCTCTC 1348  
 DB 252 ProGlyaspserGlyGlyProLeu 259  
 RESULT 23

```

EFTS_BOVIN
ID EFTS_BOVIN STANDARD; PRT; 338 AA.
AC P43896;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Elongation factor TS, mitochondrial precursor (EF-Ts) (EF-Tsmt).
GN TSM.
OS Bos taurus (Bovine).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
OC Bovidae; Bovinae; Bos.
OX NCBI_TaxID=9913;
RN [1]
RP SEQUENCE FROM N.A. AND PARTIAL SEQUENCE.
RC TISSUE=Liver;
RX MEDLINE=95340508; Pubmed=7615523;
RA Xin H., Morita V., Burkhardt W., Spremulli L.L.;
RT "Cloning and expression of mitochondrial translational elongation
RT factor is from bovine and human liver.";
RL J. Biol. Chem. 270:17243-17249(1995).
CC -1- FUNCTION: ASSOCIATES WITH THE EF-TU.GDP COMPLEX AND INDUCES THE
CC EXCHANGE OF GDP TO GTP, IT REMAINS BOUND TO THE AMINOACYL-TRNA.
CC EF-TU.GDP COMPLEX UP TO THE GTP HYDROLYSIS STAGE ON THE RIBOSOME
CC (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: Mitochondrial.
CC -1- SIMILARITY: BELONGS TO THE EF-TS FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: L37935; AAA96807.1; -
DR InterPro: IPR001816; EF_TS.
DR InterPro: IPR000449; UBA_domain.
DR Pfam: PF00627; UBA; 1.
DR TIGRFAMs: TIGR00116; tsf; 1.
DR PROSITE: PS01126; EF_TS_1; 1.
DR PROSITE: PS01127; EF_TS_2; 1.
KW Elongation factor; Protein biosynthesis; Mitochondrion;
KW Transist peptide.
FT TRANSIT 1 55 MITOCHONDRION.
FT CHAIN 56 338 ELONGATION FACTOR TS.
SQ SEQUENCE 338 AA; 36602 MW; 449236159482D068 CRC64;

Alignment Scores:
Pred. No.: 8.69 Length: 338
Score: 8.00 Matches: 8
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 1.68% Indels: 0
DB: 1 Gaps: 0

US-10-020-441-1 (1-1477) x EFTS_BOVIN (1-338)
QY 1340 GGACCTCTCCATCCATCCCTTCAA 1363
DB 26 GtYpProlLeuProSerLeuGln 33

RESULT 24
INTEB_ECOLI
ID INTEB_ECOLI STANDARD; PRT; 396 AA.
AC P39347;
DT 01-FEB-1995 (Rel. 31, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Propage P4 integrase (Int(P4)).
GN INTEB OR B4271.

```

```

OS Escherichia coli.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
OX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=K12 / MG1655;
RX MEDLINE=95334362; Pubmed=7610040;
RA Burland V.D., Plunkett G. III, Sofia H.J., Daniels D.L.,
RA Blattner F.R.;
RT "Analysis of the Escherichia coli genome VI: DNA sequence of the
RT region from 92.8 through 100 minutes.";
RL Nucleic Acids Res. 23:2105-2119(1995).
RN [2]
RP PRELIMINARY SEQUENCE OF 1-115 FROM N.A.
RC STRAIN=B;
RA Pucci M.J., Disotto L.F., Dougherty T.J.;
RL Submitted (SEP-1992) to the EMBL/GenBank/DBJ databases.
CC -1- SIMILARITY: BELONGS TO THE "PHAGE" INTEGRASE FAMILY.
CC -1- CAUTION: ADDITIONAL HOMOLOGY TO THE INTEGRASE FAMILY EXISTS
CC UPSTREAM OF THE CURRENT START CODON.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: U14003; AAA97167.1; -
DR EMBL: AE000498; AAC77227.1; -
DR EMBL: M96355; -; NOT_ANNOTATED_CDS.
DR EcoGene: EG12364; intb.
DR InterPro: IPR002104; Phage_Integrase.
DR Pfam: PF00589; Phage_Integrase; 1.
KW DNA recombination; DNA integration; Complete proteome.
FT ACT_SITE 353 353 TRANSIENT COVALENT LINKAGE TO DNA DURING
FT STRAND CLEAVAGE AND REJOINING (BY
FT SIMILARITY).
SQ SEQUENCE 396 AA; 45582 MW; DE3762C1EFC8BFCC CRC64;

Alignment Scores:
Pred. No.: 8.46 Length: 396
Score: 8.00 Matches: 8
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 1.68% Indels: 0
DB: 1 Gaps: 0

US-10-020-441-1 (1-1477) x INTEB_ECOLI (1-396)
QY 914 CACTTGAGGAGAGAGACATCATG 937
DB 359 HistLeuGluGluArgGlyLeuMet 366

RESULT 25
SKS1_YEAST
ID SKS1_YEAST STANDARD; PRT; 502 AA.
AC O12505;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Serine/threonine-protein kinase SKS1 (DC 2.7.1.-) (Suppressor kinase
DE of SNF3).
GN SKS1 OR SHA3 OR YPL026C OR LPB5.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=97103770; Pubmed=8948096;

```



RA Yang Z., Bisson L.F.;  
 RT "The SKS1 protein kinase is a multicopy suppressor of the snf3  
 RT mutation of *Saccharomyces cerevisiae*.";  
 RL Yeast 12:1407-1419(1996).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=S288C / AB972;  
 RX MEDLINE=97313273; PubMed=9169875;  
 RA Bussey H., Storms R.K., Ahmed A., Albermann K., Allen E., Ansoorge W.,  
 RA Arunjo R., Aparicio A., Barrell B.G., Badcock K., Benes V.,  
 RA Borstein D., Bowman S., Bruckner M., Carpenter J., Cherry J.M.,  
 RA Chung E., Churcher C.M., Coster F., Davis R., Davis R.W.,  
 RA Dietrich F.S., Deltus H., Dipolo T., Dubois E., Duesterhoeft A.,  
 RA Duncan M., Floeth M., Fortin N., Fritzen J.D., Fritz C., Goffeau A.,  
 RA Hall J., Heblert U., Heumann K., Hilbert H., Hillier L.,  
 RA Hunnicke-Smith S., Hyman R., Johnston M., Kalman S., Kleine K.,  
 RA Komp C., Kuril O., Lashkari D., Lew H., Lin A., Lin D., Louis E.J.,  
 RA Matthe R., Messinguy F., Mewes H.-W., Mitalipati S., Moestl D.,  
 RA Mueller-Auer S., Namath A., Neutwich U., Oefner P., Pearson D.,  
 RA Petel F.X., Pohl T.M., Purnelle D., Schafer M., Scharfe M.,  
 RA Scherens B., Schramm S., Schroeder M., Sdicu A.M., Tettelin H.,  
 RA Urrestearazu L.A., Ushinsky S., Vierendeels F., Vissers S., Voss H.,  
 RA Walsh S.V., Wamuldt R., Wang Y., Wedler E., Wedler H., Winnelt E.,  
 RA Zhong W.W., Zollner A., Vo D.H., Hahl J.;  
 RT "The nucleotide sequence of *Saccharomyces cerevisiae* chromosome XVI.";  
 RL Nature 387:103-105(1997).  
 CC -1- FUNCTION: MAY HAVE A ROLE IN GLUCOSE REGULATION.  
 CC -1- SIMILARITY: BELONGS TO THE SER/THR FAMILY OF PROTEIN KINASES.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; U30613; AAC49570.1; -;  
 DR SCD; S0005947; SKSL  
 DR InterPro: IPR000719; Euk\_Pkinase.  
 DR InterPro: IPR002290; Ser\_Thr\_Pkinase.  
 DR Pfam; PF00069; Pkinase; 1.  
 DR ProDom; PD000001; Euk\_Pkinase; 1.  
 DR PROSITE; PS00107; PROTEIN\_KINASE\_ATP; 1.  
 DR PROSITE; PS00108; PROTEIN\_KINASE\_ST; 1.  
 DR PROSITE; PS50011; PROTEIN\_KINASE\_DOM; 1.  
 KW Transferase; Serine/threonine-protein kinase; ATP-binding.  
 FT DOMAIN 10 338 PROTEIN KINASE.  
 FT NP\_BIND 16 24 ATP (BY SIMILARITY).  
 FT BINDING 39 39 ATP.  
 FT ACT\_SITE 186 186 BY SIMILARITY.  
 FT MUTAGEN 39 39 K->R: LOSS OF ACTIVITY.  
 SQ SEQUENCE 502 AA; 57844 MW; B106D084BAFA1B5 CRC64;  
 Alignment Scores:  
 Pred. No.: 8.14 Length: 502  
 Score: 8.00 Matches: 8  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 1.69% Indels: 0  
 DB: 1 Gaps: 0  
 US-10-020-441-1 (1-1477) x SKSL\_YEAST (1-502)  
 OY 1223 ACAACACACACACACACAGTACA 1200  
 Db 438 ThrAsnAsnHisAsnAsnSerThr 445  
 RESULT 26  
 ITA2\_BOVIN  
 ID ITA2\_BOVIN STANDARD; PRT; 1170 AA.  
 AC PS3710;

DT 01-OCT-1996 (Rel. 34, Created)  
 DT 01-OCT-1996 (Rel. 34, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE Integrin alpha-2 precursor (Platelet membrane glycoprotein Ia) (GP1a)  
 DE (Collagen receptor) (VLA-2 alpha chain) (CD49b) (Fragment).  
 GN ITGA2.  
 OS Bos taurus (Bovine).  
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;  
 OC Bovidae; Bovinae; Bos.  
 OX NCBI\_TaxID=9913;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=94193647; PubMed=7511592;  
 RA Kanata T., Puzon W., Takada Y.;  
 RT "Identification of putative ligand binding sites within I domain of  
 RL Integrin alpha 2 beta 1 (VLA-2, CD49b/CD29).";  
 RL J. Biol. Chem. 269:9659-9663(1994).  
 CC -1- FUNCTION: INTEGRIN ALPHA-2/BETA-1 IS A RECEPTOR FOR LAMININ.  
 CC COLLAGEN, COLLAGEN C-PROPEPTIDES, FIBRONECTIN AND E-CADHERIN. IT  
 CC RECOGNIZES THE PROLINE-HYDROXYLATED SEQUENCE G-F-P-G-E-R IN  
 CC COLLAGEN. IT IS RESPONSIBLE FOR ADHESION OF PLATELETS AND OTHER  
 CC CELLS TO COLLAGENS, MODULATION OF COLLAGEN AND COLLAGENASE GENE  
 CC EXPRESSION, FORCE GENERATION AND ORGANIZATION OF NEWLY SYNTHESIZED  
 CC EXTRACELLULAR MATRIX.  
 CC -1- SUBUNIT: HETERODIMER OF AN ALPHA AND A BETA SUBUNIT. ALPHA-2  
 CC ASSOCIATES WITH BETA-1.  
 CC -1- SUBCELLULAR LOCATION: Type I membrane protein.  
 CC -1- DOMAIN: THE INTEGRIN I-DOMAIN (INSERT) IS A VWFA DOMAIN. INTEGRINS  
 CC WITH I-DOMAINS DO NOT UNDERGO PROTEASE CLEAVAGE.  
 CC -1- SIMILARITY: BELONGS TO THE INTEGRIN ALPHA CHAIN FAMILY.  
 CC -1- SIMILARITY: CONTAINS 1 VWFA DOMAIN.  
 CC -1- SIMILARITY: CONTAINS 7 FG-GAP REPEATS.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; L25886; AAB59255.1; -;  
 DR HSSP; P17301; IAOX.  
 DR InterPro: IPR000413; Integrin\_alpha.  
 DR InterPro: IPR002035; VWF\_A.  
 DR Pfam; PF00092; vwa; 1.  
 DR Pfam; PF00357; Integrin\_A; 1.  
 DR Pfam; PF01839; FG-GAP; 5.  
 DR SMART; SMO0191; Int\_alpha; 5.  
 DR SMART; SMO0327; VWA; 1.  
 DR PROSITE; PS00242; INTEGRIN\_ALPHA; 1.  
 DR PROSITE; PS50234; VWFA; 1.  
 KW Integrin; Cell adhesion; Receptor; Glycoprotein; Transmembrane;  
 KW Platelet; Signal; Repeat; Polymorphism; Calcium; Magnesium.  
 FT NON\_TER 1 1  
 FT SIGNAL 1 18  
 FT CHAIN 19 1170 POTENTIAL.  
 FT DOMAIN 19 1121 INTEGRIN ALPHA-2.  
 FT TRANSMEM 1122 1143 EXTRACELLULAR (POTENTIAL).  
 FT DOMAIN 1144 1170 POTENTIAL.  
 FT REPEAT 34 92 CYTOPLASMIC (POTENTIAL).  
 FT REPEAT 2 2 FG-GAP 1.  
 FT REPEAT 177 367 FG-GAP 2.  
 FT REPEAT 2 2 VWFA.  
 FT REPEAT 423 475 FG-GAP 3.  
 FT REPEAT 477 538 FG-GAP 4.  
 FT REPEAT 540 599 FG-GAP 5.  
 FT REPEAT 604 656 FG-GAP 6.  
 FT REPEAT 488 496 FG-GAP 7.  
 FT CA\_BIND 488 496 POTENTIAL.  
 FT CA\_BIND 552 560 POTENTIAL.  
 FT CA\_BIND 616 624 POTENTIAL.  
 FT SITE 472 474 CELL ATTACHMENT SITE (POTENTIAL).  
 FT SITE 474 474

```
FT SITE 1146 1150 GFFKR MOTIF.
FT DISULFID 72 81 BY SIMILARITY.
FT DISULFID 669 726 BY SIMILARITY.
FT DISULFID 778 784 BY SIMILARITY.
FT DISULFID 854 865 BY SIMILARITY.
FT DISULFID 1008 1039 BY SIMILARITY.
FT DISULFID 1044 1049 BY SIMILARITY.
FT CARBOHYD 94 94 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 101 101 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 332 332 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 421 421 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 449 449 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 464 464 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 688 688 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 748 748 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 945 945 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1063 1063 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1070 1070 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT VARIANT 580 580 G -> V.
FT VARIANT 588 588 R -> K.
FT VARIANT 725 725 R -> S.
SQ SEQUENCE 1170 AA; 128929 MW; EECCEFC5F2448FB1 CRC64;
```

```
Alignment Scores:
Pred. No.: 7.07 Length: 1170
Score: 8.00 Matches: 8
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 1.68% Indels: 0
DB: NCBI_Taxid=9796; Gaps: 0
```

US-10-020-441-1 (1-1477) x ITA2\_BOVIN (1-1170)

QY 792 CAGTTCAGTACAGAGTGTTC 815  
Db 695 GlnpheserArValIleSer 702

```
RESULT 27
Y795_TREPA STANDARD: PRT; 52 AA.
ID Y795_TREPA
AC 083773;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical protein TP0795.
GN TP0795.
OS Treponema pallidum.
OC Bacteria; Spirochaetales; Spirochaetaceae; Treponema.
ON NCBI_Taxid=160;
RX [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Nichols;
RC MEDLINE=98332770; PubMed=9665876;
RA Fraser C.M., Norris S.J., Weinstein G.M., White O., Sutton G.G.,
RA Dodson R., Gwin M., Hickey E.K., Clayton R., Ketchum K.A.,
RA Sodergren E., Hardham J.M., McLeod M.P., Salberg S., Peterson J.,
RA Khalak H., Richardson D., Howell J.K., Chidambaram M., Uterback F.,
RA McDonald L., Arlisch P., Bowman C., Cotton M.D., Fujii C., Garland S.,
RA Hatch B., Horst K., Roberts K., Sandusky M., Weidman J., Smith H.O.,
RA Venter J.C.;
RT "Complete genome sequence of Treponema pallidum, the syphilis
spirochete."
RL Science 281:375-388(1998).
```

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL; AEO01250; AAC65767.1; -.

DR TIGR: TP0795; -  
KW Hypothetical protein; Complete proteome.  
SQ SEQUENCE 52 AA; 5770 MW; 5654464CADB69233 CRC64;

```
Alignment Scores:
Pred. No.: 143 Length: 52
Score: 7.00 Matches: 7
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 1.47% Indels: 0
DB: NCBI_Taxid=9796; Gaps: 0
```

US-10-020-441-1 (1-1477) x Y795\_TREPA (1-52)

QY 109 TCACGACGAGCAGTACTCACA 129  
Db 45 SerThrArgAlaValLeuThr 51

```
RESULT 28
EL2B_HORSE STANDARD: PRT; 73 AA.
ID EL2B_HORSE
AC P37358;
DT 01-OCT-1994 (Rel. 30, Created)
DT 01-OCT-1994 (Rel. 30, Last sequence update)
DT 01-OCT-1996 (Rel. 34, Last annotation update)
DE Neutrophil elastase 2B (EC 3.4.21.-) (Proteinase 2B) (Fragments).
OS Equus caballus (Horse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Perissodactyla; Equidae; Equus.
OX NCBI_Taxid=9796;
RX [1]
RP SEQUENCE.
RC TISSUE=Neutrophils;
RC MEDLINE=94271153; PubMed=7516152;
RA Dublin A., Potempa J., Travis J.;
RT "Structural and functional characterization of elastases from horse
RT neutrophils."
RL Biochem. J. 300:401-406(1994).
```

CC -I- FUNCTION: MAY BE INVOLVED IN THE DEGRADATION OF CONNECTIVE  
CC TISSUE IN CHRONIC LUNG DISEASE.  
CC -I- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1. ELASTASE SUBFAMILY.  
DR PIR: S44462; S44462.  
DR HSSP: P00766; 1GCD.  
DR MEROPS: S01.131; -.  
DR InterPro: IPR001254; Ser\_protease\_Try.  
DR PROSITE: PS50240; TRYPSIN\_DOM; PARTIAL.  
DR PROSITE: PS00134; TRYPSIN\_HIS; PARTIAL.  
DR PROSITE: PS00135; TRYPSIN\_SER; 1.  
KW Hydrolase; Serine protease.  
FT NON\_CONS 31 32  
FT NON\_CONS 56 57  
FT ACT\_SITE 64 64  
SQ SEQUENCE 73 AA; 7615 MW; AFFB0B330DB69041 CRC64;

```
Alignment Scores:
Pred. No.: 135 Length: 73
Score: 7.00 Matches: 7
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 1.47% Indels: 0
DB: NCBI_Taxid=9796; Gaps: 0
```

US-10-020-441-1 (1-1477) x EL2B\_HORSE (1-73)

QY 1328 GGTGACAGTGTGACCTCTC 1348  
Db 62 GlyAspSerGlyGlyProLeu 68

```
RESULT 29
Q300_MOUSE STANDARD: PRT; 77 AA.
ID Q300_MOUSE
AC 002722;
DT 01-JUN-1994 (Rel. 29, Created)
```

```

DT 01-JUN-1994 (Rel. 29, last sequence update)
DT 01-NOV-1997 (Rel. 35, last annotation update)
DE Protein Q300.
GN HPVC2 OR Q300.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Kidney;
RX MEDLINE=91237845; PubMed=1851876;
RA Wagner S., Cullman G., Knippers R.;
RT "The Q300 gene: a novel transcription unit induced in simian virus
RT 40-infected and -transformed mouse cells.";
RL J. Virol. 65:3259-3267(1991).
CC -!- FUNCTION: MIGHT PLAY A ROLE IN IMMORTALIZATION OF SV40-INFECTED
CC CELLS. MIGHT BE THE CELLULAR COUNTERPART OF THE VIRAL E5 PROTEIN.
CC -!- SUBCELLULAR LOCATION: Membrane-bound (Potential).
CC -!- SIMILARITY: TO THE HUMAN PAPILLOMAVIRUS E5 PROTEIN.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
CC
DR EMBL: X52164; CAA36417.1; -.
DR MGD: MGI:109158; HPVC2.
KW Transmembrane; Repeat.
FT DOMAIN 27 7 X 2 AA TANDEM REPEATS OF V-C.
FT TRANSMEM 27 40 POTENTIAL.
SQ SEQUENCE 77 AA; 8493 MW; 4f62a52dc066643 CRC64;

Alignment Scores:
Pred. No.: 134 Length: 77
Score: 7.00 Matches: 7
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 1.47% Indels: 0
Gaps: 0

US-10-020-441-1 (1-1477) x Q300_MOUSE (1-77)
QY 241 TGCCTTGTGTTGTGTATGC 261
Db 28 CysValcysValcysValcys 34

RESULT 30
EL2A_HORSE STANDARD; PRT; 85 AA.
ID EL2A_HORSE
AC P37357;
DT 01-OCT-1994 (Rel. 30, Created)
DT 01-OCT-1994 (Rel. 30, last sequence update)
DT 01-OCT-1996 (Rel. 34, last annotation update)
DE Neutrophil elastase 2A (EC 3.4.21.-) (Proteinase 2A) (Fragments).
OS Equus caballus (Horse).
OC Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Perissodactyla; Equidae; Equus.
OX NCBI_TaxID=9796;
RN [1]
RP SEQUENCE.
RC TISSUE=Neutrophils;
RX MEDLINE=94271153; PubMed=7516152;
RA Dublin A., Potempa J., Travis J.;
RT "Structural and functional characterization of elastases from horse
RT neutrophils.";
RL Biochem. J. 300:401-406(1994).
CC -!- FUNCTION: MAY BE INVOLVED IN THE DEGRADATION OF CONNECTIVE
CC TISSUE IN CHRONIC LUNG DISEASE.
CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1. ELASTASE SUBFAMILY.

```

```

DR PIR: S44461; S44461.
DR HSSP: P24158; 1FUJ.
DR MEROPS: S01.131; -.
DR InterPro: IPR001254; Ser_protease_Try.
DR PROSITE: PS00240; TRYPSIN_DOM; PARTIAL.
DR PROSITE: PS00134; TRYPSIN_HIS; PARTIAL.
DR PROSITE: PS00135; TRYPSIN_SER; 1.
KW Hydrolase; Serine protease.
FT NON_CONS 34 35
FT NON_CONS 59 60
FT ACT_SITE 67 67
SQ SEQUENCE 85 AA; 8891 MW; 0ABE7D44A1B90E26 CRC64;

Alignment Scores:
Pred. No.: 131 Length: 85
Score: 7.00 Matches: 7
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 1.47% Indels: 0
Gaps: 0

US-10-020-441-1 (1-1477) x EL2A_HORSE (1-85)
QY 1328 GGTGACAGTGTGACACTCTC 1348
Db 65 GlyaspserGlyglyProleu 71

RESULT 31
OL7H_MOUSE STANDARD; PRT; 111 AA.
ID OL7H_MOUSE
AC Q60893;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, last sequence update)
DT 15-JUL-1998 (Rel. 36, last annotation update)
DE Olfactory receptor 7H (M71) (Fragment).
OS OLF7.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=C57BL/6J;
RX MEDLINE=96149403; PubMed=8570653;
RA Sullivan S.L., Adamson M.C., Resler K.J., Kozak C.A., Buck L.B.;
RT "The chromosomal distribution of mouse odorant receptor genes.";
RL Proc. Natl. Acad. Sci. U.S.A. 93:884-888(1996).
CC -!- FUNCTION: PUTATIVE ODORANT RECEPTOR.
CC -!- SUBCELLULAR LOCATION: Integral membrane protein.
CC -!- MISCELLANEOUS: OLFACTORY RECEPTORS 7A-7I ARE PRODUCED BY EIGHT
CC DIFFERENT GENES WITHIN THE OLF7 COMPLEX.
CC -!- SIMILARITY: BELONGS TO FAMILY 1 OF G-PROTEIN COUPLED RECEPTORS.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
CC
DR EMBL: U28782; AAC52405.1; -.
DR MGD: MGI:104712; Olf7.
DR InterPro: IPR000276; GPCR_Rhodops.
DR PROSITE: PS00237; G_PROTEIN_RECPT_F1_1; PARTIAL.
DR PROSITE: PS00262; G_PROTEIN_RECPT_F1_2; 1.
KW G-protein coupled receptor; Transmembrane; Multigene family;
KW Olfaction.
FT NON_TER 1 1
FT DOMAIN <1 12 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 13 33 4 (POTENTIAL).
FT DOMAIN 34 72 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 73 94 5 (POTENTIAL).

```

```

FT DOMAIN 95 108 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 109 >111 6 (POTENTIAL).
FT NON_TER 111 111
SQ SEQUENCE 111 AA; 12118 MM; 643DAPFD4BB877D3 CRC64;

Alignment Scores:
Pred. No.: 126 Length: 111
Score: 7.00 Matches: 7
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 1.47% Indels: 0
DB: 1 Gaps: 0

US-10-020-441-1 (1-1477) x OL7H_MOUSE (1-111)

QY 251 TTGTGTATGACAGTGGTTTG 271
   |||||
Db 20 PheValTyralValGlyLeu 26

RESULT 32
PSC2_RAT STANDARD; PRT; 112 AA.
AC P02781;
DT 21-JUL-1986 (Rel. 01, Created)
DT 01-JAN-1988 (Rel. 06, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE Prostatic steroid-binding protein C2 chain precursor (Prostatein
DE peptide C2).
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RA MEDLINE=87146484; PubMed=2881277;
RA Delaey B., Dirckx L., Decourt J.-L., Claessens F., Peeters B.,
RA Rombaerts W.;
RT "Rat prostatic binding protein: the complete sequence of the C2 gene
RT and its flanking regions.";
RL Nucleic Acids Res. 15:1627-1641(1987).
RN [2]
RP SEQUENCE OF 21-112.
RA MEDLINE=83209619; PubMed=6343081;
RA Peeters B., Heyns W., Mous J., Rombaerts W.;
RT "Structural studies on rat prostatic binding protein. The primary
RT structure of component C2 from subunit S.";
RL Eur. J. Biochem. 132:669-679(1983).
RN [3]
RP SEQUENCE OF 1-100 FROM N.A.
RA MEDLINE=82220075; PubMed=6896362;
RA Parker M., Needham M., White R.;
RT "Prostatic steroid binding protein: gene duplication and steroid
RT binding.";
RL Nature 298:92-94(1982).
CC -1- FUNCTION: PART OF PROSTATEIN WHICH IS THE MAJOR SECRETORY
CC GLYCOPROTEIN OF VENTRAL PROSTATE GLAND.
CC -1- SUBUNIT: PROSTATEIN IS COMPOSED OF THREE DIFFERENT PEPTIDES CALLED
CC C1, C2 AND C3. THESE FORM COVALENT C1:C3 (F) AND C2:C3 (S)
CC HETERODIMERS WHOSE NONCOVALENT ASSOCIATION FORMS TETRAMERIC
CC (C1:C3/C3:C2) PROSTATEIN MOLECULES.
CC -1- SUBCELLULAR LOCATION: Secreted.
CC -1- PTM: LINKED BY THREE DISULFIDE BONDS TO C3.
CC -1- MISCELLANEOUS: THE HETERODIMER CAN BIND NON-POLAR STEROIDS,
CC CHOLESTEROL AND A GROUP OF SMALL PROLINE-RICH PEPTIDES.
CC -1- SIMILARITY: BELONGS TO THE UTEROGLOBIN FAMILY. LIPOPHILIN
CC SUBFAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/

```

```

CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: X05034; CAA28708.1; -.
DR EMBL: V01256; CAA24569.1; -.
DR EMBL: J00776; AAA51641.1; -.
DR PIR: A03251; BORT2.
DR PIR: A26671; A26671.
DR InterPro: IPR000329; Uterogloblin.
DR PROSITE: PS00403; UTEROGLOBIN_1; FALSE_NEG.
DR PROSITE: PS00404; UTEROGLOBIN_2; FALSE_NEG.
KW Steroid-binding; Signal.
FT SIGNAL 1 20
FT CHAIN 21 112 PROSTATIC STEROID-BINDING PROTEIN C2
FT MOD_RES 21 21 BLOCKED.
FT DISULFID 28 28 INTERCHAIN (WITH C3) (PROBABLE).
FT DISULFID 69 69 INTERCHAIN (WITH C3) (PROBABLE).
FT DISULFID 92 92 INTERCHAIN (WITH C3) (PROBABLE).
FT CONFLICT 26 26 MISSING (IN REF. 3).
FT CONFLICT 88 88 I -> T (IN REF. 3).
FT CONFLICT 96 112 VWLIQNEPPRRWFSEIN -> YGYK (IN REF. 3).
SQ SEQUENCE 112 AA; 12828 MM; DA65A6A82E677864 CRC64;

Alignment Scores:
Pred. No.: 126 Length: 112
Score: 7.00 Matches: 7
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 1.47% Indels: 0
DB: 1 Gaps: 0

US-10-020-441-1 (1-1477) x PSC2_RAT (1-112)

QY 135 TCATTGTGTTGCTCACCATTT 155
   |||||
Db 4 SerLeuGlySerLeuThrIle 10

RESULT 33
RS13_ECOLI STANDARD; PRT; 117 AA.
ID RS13_ECOLI
AC P02369;
DT 21-JUL-1986 (Rel. 01, Created)
DT 21-JUL-1986 (Rel. 01, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE 30S ribosomal protein S13.
GN RPSM OR B3298 OR Z4668 OR ECS4163.
OS Escherichia coli, and
OS Escherichia coli O157:H7.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
OX NCBI_TaxID=562, 83334;
RN [1]
RP SEQUENCE FROM N.A.
RA STRAIN=K12;
RC MEDLINE=85242076; PubMed=2989779;
RA Bedwell D.M., Davis G.R., Gosink M., Post L.E., Nomura M., Kestler H.,
RA Zengel J.M., Lindahl L.;
RT "Nucleotide sequence of the alpha ribosomal protein operon of
RT Escherichia coli.";
RL Nucleic Acids Res. 13:3891-3903(1985).
RN [2]
RP SEQUENCE FROM N.A.
RA STRAIN=K12 / MG1655;
RC MEDLINE=97426617; PubMed=9278503;
RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA Mau B., Shao Y.;
RT "The complete genome sequence of Escherichia coli K-12.";
RN [3]
RP SEQUENCE FROM N.A.
RA STRAIN=O157:H7 / EDL933 / ATCC 700927;

```

RX MEDLINE=21074935; PubMed=11206551;  
 RA Perna N.T., Plunkett G., Ili, Burland V., Mau B., Glasner J.D.,  
 RA Rose D.J., Mayhew G.F., Evans P.S., Gregor J., Kirkpatrick H.A.,  
 RA Posfai G., Hackett J., Klink S., Boutin A., Shao Y., Miller L.,  
 RA Grobbeck E.U., Davis N.W., Lim A., Dimalanta E.T., Potamoukis K.,  
 RA Apodaca J., Anantharaman T.S., Lin J., Yen G., Schwartz D.C.,  
 RA Welch R.A., Blattner F.R.;  
 RT "Genome sequence of enterohaemorrhagic *Escherichia coli* O157:H7";  
 RT Nature 409:529-533(2001).  
 RL  
 RN  
 RP  
 RC  
 RX STRAIN=O157:H7 / RIMD 0509952;  
 RX MEDLINE=21156231; PubMed=11258796;  
 RA Hayashi T., Makino K., Ohnishi M., Kurokawa K., Ishii K., Yokoyama K.,  
 RA Han C.-G., Ohtsubo E., Nakayama K., Murata T., Tanaka M., Tobe T.,  
 RA Iida T., Takami H., Honda T., Sasakawa C., Ogasawara N., Yasunaga T.,  
 RA Kunara S., Shiba T., Hattori M., Shinagawa H.;  
 RT "Complete genome sequence of enterohaemorrhagic *Escherichia coli*  
 RT O157:H7 and genomic comparison with a laboratory strain K-12";  
 RL DNA Res. 8:11-22(2001).  
 RN  
 RP  
 RC  
 RX STRAIN=K;  
 RX MEDLINE=77248097; PubMed=330375;  
 RA Lindemann H., Wittmann-Liebold B.;  
 RT "Primary structure of protein S13 from the small subunit of  
 RT *Escherichia coli* ribosomes";  
 RL Hoppe-Seyler's Z. Physiol. Chem. 358:843-863(1977).  
 RN  
 RP  
 RC  
 RX MEDLINE=80182128; PubMed=6154696;  
 RA Post L.E., Arfsten A.E., Davis G.R., Nomura M.;  
 RT "DNA sequence of the promoter region for the alpha ribosomal protein  
 RT operon in *Escherichia coli*";  
 RL J. Biol. Chem. 255:4653-4659(1980).  
 RN  
 RP  
 RC  
 RX MEDLINE=82025589; PubMed=6793240;  
 RA Mura A., Krueger J.H., Itoh S., de Boer H.A., Nomura M.;  
 RT "Growth-rate-dependent regulation of ribosome synthesis in *E. coli*;  
 RT expression of the *lacZ* and *galK* genes fused to ribosomal promoters";  
 RL Cell 25:773-782(1981).  
 RN  
 RP  
 RC  
 RX MEDLINE=99196679; PubMed=10094780;  
 RA Arnold R.J., Reilly J.P.;  
 RT "Observation of *Escherichia coli* ribosomal proteins and their  
 RT posttranslational modifications by mass spectrometry";  
 RL Anal. Biochem. 269:105-112(1999).  
 CC  
 CC -1- FUNCTION: INVOLVED IN THE BINDING OF FMET-TRNA AND, HENCE, IN THE  
 CC INITIATION OF TRANSLATION  
 CC  
 CC -1- MASS SPECTROMETRY: MW=12968.1; METHOD=MALDI.  
 CC  
 CC -1- SIMILARITY: BELONGS TO THE S13P FAMILY OF RIBOSOMAL PROTEINS.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC -----  
 DR EMBL: X02543; CAA26392.1; -;  
 DR EMBL: U18997; AAB58093.1; -;  
 DR EMBL: AE000407; AAC76323.1; -;  
 DR EMBL: AE005556; AAC58419.1; -;  
 DR EMBL: AP002564; BAB37586.1; -;  
 DR EMBL: M12432; AAB83903.1; -;  
 DR EMBL: M10213; AAA72457.1; -;  
 DR PIR: A23807; R3EC13.  
 DR ECGene; EG10912; TFSM.  
 DR InterPro; IPR001892; Ribosomal\_S13.  
 DR Pfam; PF00416; Ribosomal\_S13; 1.

DR ProDom: PD001363; Ribosomal\_S13; 1.  
 DR PROSITE: PS00646; RIBOSOMAL\_S13; 1.  
 KW Ribosomal protein; Complete proteome.  
 FT INT\_MET 0  
 SQ SEQUENCE 117 AA; 12968 MW; 3277C328BBD0D3D9 CRC64;  
 Alignment Scores:  
 Pred. No.: 125 Length: 117  
 Score: 7.00 Matches: 7  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 1.48% Indels: 0  
 DB: 1 Gaps: 0  
 US-10-020-441-1 (1-1477) x RS13\_ECOLI (1-117)  
 QY 249 ACAACGACGACGACGACGAAA 229  
 DB 103 ThrAsnAlaArgThrArgLys 109  
 RSURF\_34  
 ID RS13\_SHEVI STANDARD; PRT; 118 AA.  
 AC Q950R1;  
 DT 15-JUN-2002 (Rel. 41, Created)  
 DT 15-JUN-2002 (Rel. 41, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE 30S ribosomal protein S13.  
 GN RPSM.  
 OS *Shewanella violacea*.  
 OC Bacteria; Proteobacteria; gamma subdivision; Alteromonadaceae;  
 CC *Shewanella*.  
 OX NCBI\_TaxID=60217;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=DSS12;  
 RX PubMed=11111034;  
 RA Nakasone K., Ikegami A., Fujii S., Kato C., Horikoshi K.;  
 RT "Isolation and piezoresponse of the *rpoA* gene encoding the RNA  
 RT polymerase alpha subunit from the deep-sea piezophilic bacterium  
 RT *Shewanella violacea*";  
 RL FEWS Microbiol. Lett. 193:261-268(2000).  
 CC  
 CC -1- FUNCTION: INVOLVED IN THE BINDING OF FMET-TRNA AND, HENCE, IN THE  
 CC INITIATION OF TRANSLATION (BY SIMILARITY).  
 CC  
 CC -1- SIMILARITY: BELONGS TO THE S13P FAMILY OF RIBOSOMAL PROTEINS.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC -----  
 DR EMBL: AB032408; BAA84522.1; -;  
 DR InterPro: IPR001892; Ribosomal\_S13.  
 DR Pfam: PF00416; Ribosomal\_S13; 1.  
 DR PROSITE: PD001363; Ribosomal\_S13; 1.  
 DR PROSITE: PS00646; RIBOSOMAL\_S13; 1.  
 KW Ribosomal protein.  
 SQ SEQUENCE 118 AA; 13277 MW; C1BD51545CFB1204 CRC64;  
 Alignment Scores:  
 Pred. No.: 125 Length: 118  
 Score: 7.00 Matches: 7  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 1.48% Indels: 0  
 DB: 1 Gaps: 0  
 US-10-020-441-1 (1-1477) x RS13\_SHEVI (1-118)  
 QY 249 ACAACGACGACGACGACGAAA 229



FT SIGNAL 1 21 POTENTIAL.  
FT CHAIN 22 120 ALPHA-2 MATING PHEROMONE.  
FT PEPTIDE 87 99 MATING FACTOR ALPHA (1ST COPY).  
FT PEPTIDE 108 120 MATING FACTOR ALPHA (2ND COPY).  
SQ SEQUENCE 120 AA; 13271 MW; 10BF3FDB985FBBD2 CRC64;

## Alignment Scores:

Pred. No.: 124 Length: 120  
Score: 7.00 Matches: 7  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 1.48% Indels: 0  
DB: 1 Gaps: 0

US-10-020-441-1 (1-1477) x MFA4\_YEAST (1-120)

QY 1176 GAGAAAGACAAACACACG 1156

Db 70 GtutysgiuglinsrthThr 76

## RESULT 37

RS13\_HAEIN STANDARD; PRT; 121 AA.

ID RS13\_HAEIN

AC P44380;

DT 01-NOV-1995 (Rel. 32, Created)

DT 01-NOV-1995 (Rel. 32, Last sequence update)

DT 15-JUN-2002 (Rel. 41, Last annotation update)

DE 30S ribosomal protein S13.

GN RPSM OR RPS13 OR H10799.

OS Haemophilus influenzae.

OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;

CC Haemophilus

OX NCBI\_TaxID=727;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=Rd / KW20 / ATCC 51907;

RX MEDLINE=95350630; PubMed=7542800;

RA Fleischmann R.D., Adams M.D., White O., Clayton R.A., Kirkness E.F.,

RA Kelleraage A.R., Bult C.J., Tomb J.-F., Dougherty B.A., Merrick J.M.,

RA McEnney K., Sutton G., Fitzhugh W., Fields C.A., Gocayne J.D.,

RA Scott J.D., Shirley R., Liu L.-I., Glodek A., Kelley J.M.,

RA Weidman J.F., Phillips C.A., Spriggs T., Hedblom E., Cotton M.D.,

RA Ulfenback T.R., Hanna M.C., Nguyen P.T., Saudke D.M., Brandon R.C.,

RA Flute L.D., Fitchman J.L., Fuhrmann J.L., Geoghagen N.S.M.,

RA Gnehm C.L., McDonald L.A., Small K.V., Fraser C.M., Smith H.O.,

RA Venter J.C.;

RT "Whole-genome random sequencing and assembly of Haemophilus influenzae

RD Rd.";

RL Science 269:496-512(1995).

CC -1- FUNCTION: INVOLVED IN THE BINDING OF FMET-tRNA AND, HENCE, IN THE

CC INITIATION OF TRANSLATION (BY SIMILARITY).

CC -1- SIMILARITY: BELONGS TO THE S13P FAMILY OF RIBOSOMAL PROTEINS.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC EMBL: U32762; AAC22458.1; -.

DR TIGR: H10799; -.

DR InterPro: IPR001892; Ribosomal\_S13.

DR Pfam: PF00416; Ribosomal\_S13; 1.

DR ProDom: PD001363; Ribosomal\_S13; 1.

DR PROSITE: PS00646; RIBOSOMAL\_S13; 1.

KW Ribosomal protein; Complete proteome.

FT INT\_MET 0 By SIMILARITY

SQ SEQUENCE 121 AA; 13565 MW; 48463C07BF70E9A8 CRC64;

Alignment Scores: 124 Length: 121  
Pred. No.: 124

Score: 7.00 Matches: 7  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 1.48% Indels: 0  
DB: 1 Gaps: 0

US-10-020-441-1 (1-1477) x RS13\_HAEIN (1-121)

QY 249 ACAACGACGACGACGACGCAA 229

Db 103 ThranaAlaArgThrArgLys 109

## RESULT 38

RS13\_TREPA

ID RS13\_TREPA

AC O83240;

DT 15-DEC-1998 (Rel. 37, Created)

DT 15-DEC-1998 (Rel. 37, Last sequence update)

DT 16-OCT-2001 (Rel. 40, Last annotation update)

DE 30S ribosomal protein S13.

GN RPSM OR TP0210.

OS Treponema pallidum.

OC Bacteria; Spirochaetales; Spirochaetaceae; Treponema.

CC Treponema

OX NCBI\_TaxID=160;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=Nichols;

RX MEDLINE=98332770; PubMed=9665876;

RA Fraser C.M., Norris S.J., Weinstein G.M., White O., Sutton G.G.,

RA Dodson R., Gwin M., Hickey E.K., Clayton R., Ketchum K.A.,

RA Sodergren E., Hardham J.M., McLeod M.P., Salzberg S., Peterson J.,

RA Khalak H., Richardson D., Howell J.K., Chidambaram M., Ulfenback T.,

RA McDonald L., Artach P., Bowman C., Cotton M.D., Fujii C., Garland S.,

RA Hatch B., Horst K., Roberts K., Sandusky M., Weidman J., Smith H.O.,

RA Venter J.C.;

RT "Complete genome sequence of Treponema pallidum, the syphilis

RD spirochete."

RL Science 281:375-388(1998).

CC -1- FUNCTION: INVOLVED IN THE BINDING OF FMET-tRNA AND, HENCE, IN THE

CC INITIATION OF TRANSLATION (BY SIMILARITY).

CC -1- SIMILARITY: BELONGS TO THE S13P FAMILY OF RIBOSOMAL PROTEINS.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC EMBL: AE001203; AAC65199.1; -.

DR TIGR: TP0210; -.

DR InterPro: IPR001892; Ribosomal\_S13.

DR Pfam: PF00416; Ribosomal\_S13; 1.

DR ProDom: PD001363; Ribosomal\_S13; 1.

DR PROSITE: PS00646; RIBOSOMAL\_S13; 1.

KW Ribosomal protein; Complete proteome.

SQ SEQUENCE 121 AA; 13634 MW; 609620F40CA059B CRC64;

Alignment Scores: 124 Length: 121  
Pred. No.: 124

Score: 7.00 Matches: 7

Percent Similarity: 100.00% Conservative: 0

Best Local Similarity: 100.00% Mismatches: 0

Query Match: 1.48% Indels: 0

DB: 1 Gaps: 0

US-10-020-441-1 (1-1477) x RS13\_TREPA (1-121)

QY 249 ACAACGACGACGACGACGCAA 229

Db 104 ThranaAlaArgThrArgLys 110





GenCore version 5.1.4.p5.4578  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - protein search, using frame\_plus.n2p model

Run on: April 10, 2003, 07:47:30 ; Search time 128 seconds  
(without alignments)  
4755.181 Million cell updates/sec

Title: US-10-020-441-1  
Perfect score: 477  
Sequence: 1 ggtcacactggttgatagc.....ttgtacgtcccaatttga 1477

Scoring table:  
OLIGO  
Xgapop 60.0 , Xgapext 60.0  
Ygapop 60.0 , Ygapext 60.0  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 671580 segs, 206047115 residues

Word size: 1

Total number of hits satisfying chosen parameters: 1343044

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Listing first 1000 summaries

Command line parameters:

-MODEL-frame+ .n2p.model -DEV-xlp  
-O/cg92.1/USPTO.spool/US10020441/runat\_10042003\_074720\_25909/app\_query.fasta\_1.1671  
-DB-SPTREMBL\_21 -QFMT-fastan -SUFFIX-oligo.rspt -MINMATCH=0.1 -LOOPEXT=0  
-LOOPEXT=0 -UNITS-bits -START=1 -END=1 -MATRIX-oligo -TRANS-human40.cdi  
-LIST=1000 -DOCALLIGN=200 -THR SCORE=quality -THR\_MIN=1 -ALIGN=40 -MODE=LOCAL  
-OUTFMT-pico -NOR-ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=2000000000  
-USER=US10020441@cgn.1.172.8runat\_10042003\_074720\_25909 -NCP=6 -ICPU=3  
-NO\_XLPXY -NO\_MMAP -LARGEQUERY -NEG\_SCORES=0 -WAIT -LONGLOG -DEV.TIMEOUT=120  
-WARN.TIMEOUT=30 -THREADS=1 -XGAPOP=60 -XGAPEXT=60 -Fgapop=6 -Fgapext=7  
-YGAPOP=60 -YGAPEXT=60 -DELOP=6 -DELEXT=7

Database :

SPTREMBL\_21:\*  
1: sp\_archaea:\*  
2: sp\_bacteria:\*  
3: sp\_fungi:\*  
4: sp\_human:\*  
5: sp\_invertebrate:\*  
6: sp\_mammal:\*  
7: sp\_mhc:\*  
8: sp\_organelle:\*  
9: sp\_phage:\*  
10: sp\_plant:\*  
11: sp\_rodent:\*  
12: sp\_virus:\*  
13: sp\_vertebrate:\*  
14: sp\_unclassified:\*  
15: sp\_virus:\*  
16: sp\_bacteriap:\*  
17: sp\_archaeap:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Length	ID	Description
1	41	8.6	274 5	Q26552 schistosoma

2	23	4.8	257 5	Q26553 schistosoma
3	12	2.5	44 11	Q9Kf9 mus musculus
4	10	2.1	188 5	Q9N859 trypanosoma
5	10	2.1	361 5	Q94176 caenorhabdi
6	9	1.9	104 5	Q9NLE3 leishmania
7	9	1.9	110 16	Q8VNO0 anabaena sp
8	9	1.9	400 4	Q8WYX9 homo sapien
9	9	1.9	696 2	Q07651 campylobact
10	9	1.9	696 16	Q9PPG3 campylobact
11	8	1.7	46 6	Q9TTC4 ateles belz
12	8	1.7	105 10	Q9A0Y9 oryza sativ
13	8	1.7	115 4	Q9UDH5 gduh5 homo sapien
14	8	1.7	120 3	Q96UD9 neurospora
15	8	1.7	128 11	Q9EQ13 mus musculus
16	8	1.7	132 12	Q9W335 mus musculus
17	8	1.7	156 16	Q92GJ4 rickettsia
18	8	1.7	176 11	Q8VDE7 aedes albop
19	8	1.7	214 5	Q9XVW2 schistosoma
20	8	1.7	223 5	Q9XY58 ctenocephal
21	8	1.7	223 5	Q9VBY4 drosophila
22	8	1.7	232 5	Q9TXF6 bombyx mori
23	8	1.7	235 5	Q9GTX7 aedes albop
24	8	1.7	236 5	Q9TYH3 schistosoma
25	8	1.7	239 11	Q9EQT2 mus musculus
26	8	1.7	242 8	Q8WL74 mionandra c
27	8	1.7	246 11	Q9EPD0 mus musculus
28	8	1.7	246 11	Q9EPD9 mus musculus
29	8	1.7	246 11	Q9RC8 ratius norv
30	8	1.7	246 11	Q91VB1 mus musculus
31	8	1.7	247 5	Q17039 anopheles g
32	8	1.7	247 11	Q8WL73 cordoba ar
33	8	1.7	247 11	Q08732 mesocricetu
34	8	1.7	247 11	Q70164 mesocricetu
35	8	1.7	247 11	Q35342 mesocricetu
36	8	1.7	247 11	Q70500 ratius norv
37	8	1.7	255 13	Q34289 salvelinus
38	8	1.7	264 2	Q54426 xanthomonas
39	8	1.7	265 5	Q9VAX6 drosophila
40	8	1.7	268 5	Q17030 anopheles g
41	8	1.7	272 5	Q18459 heterodera
42	8	1.7	273 8	Q94U06 leishmania
43	8	1.7	275 2	Q53983 myxococcus
44	8	1.7	278 16	Q82E64 yersinia pe
45	8	1.7	281 5	Q76898 drosophila
46	8	1.7	301 2	Q53756 streptomyce
47	8	1.7	316 5	Q9VAG2 drosophila
48	8	1.7	334 5	Q9VF28 drosophila
49	8	1.7	347 5	Q18457 heterodera
50	8	1.7	365 5	Q97366 holotrichia
51	8	1.7	367 5	Q9VFW0 drosophila
52	8	1.7	390 5	Q9V157 drosophila
53	8	1.7	393 5	Q9W3D7 drosophila
54	8	1.7	395 2	Q9AL66 shigella fl
55	8	1.7	396 2	Q8VNO8 escherichia
56	8	1.7	396 5	Q9VW19 drosophila
57	8	1.7	408 5	Q9U320 caenorhabdi
58	8	1.7	415 5	Q9ADX1 escherichia
59	8	1.7	421 2	Q8VON9 escherichia
60	8	1.7	421 2	Q8VLM4 escherichia
61	8	1.7	421 16	Q8XCS1 escherichia
62	8	1.7	436 16	Q9XAZ7 streptomyce
63	8	1.7	441 5	Q9XVX0 bombyx mori
64	8	1.7	488 5	Q9TYH4 schistosoma
65	8	1.7	604 8	Q8SFB5 haematopus
66	8	1.7	631 10	Q40156 lycopersico
67	8	1.7	667 5	Q9BJM1 trichinella
68	8	1.7	733 5	Q9VFX9 drosophila
69	8	1.7	745 13	Q91701 xenopus lae
70	8	1.7	747 13	Q91900 xenopus lae
71	8	1.7	888 10	Q8W0U4 sorghum b1c
72	8	1.7	891 5	Q18602 caenorhabdi
73	8	1.7	1004 13	P79953 xenopus lae
74	8	1.7	1145 5	Q965N2 caenorhabdi

















Query Match: 4.82% Indels: 0  
DB: 5 Gaps: 0  
US-10-020-441-1 (1-1477) x Q26553 (1-257)

OY 316 TTCTCAGTACAGTGGACCAACAGCATCCATCACCACCTGAGATTAG 375  
|  
DB 82 PheuthrileuArganglyAspGlnGlnIleHSHSGInProserGlyVallys 101  
|  
OY 376 GTGCACCA 384  
|  
DB 102 ValAlaPro 104

RESULT 3  
O99KF9 PRELIMINARY: PRT: 44 AA.  
ID O99KF9  
AC O99KF9  
DT 01-JUN-2001 (TREMBlrel. 17, Created)  
DT 01-JUN-2001 (TREMBlrel. 17, Last sequence update)  
DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)  
DE Hypothetical 4.9 kDa protein.  
GN 3110001K13RIK.  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
OX NCBI\_TaxID=10090;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Strausberg R.;  
RL Submitted (MAR-2001) to the EMBL/GenBank/DBJ databases.  
DR EMBL: BC004681; AAO04681.1; -  
DR MGD: MGI:1919666; 3110001K13RIK.  
KM Hypothetical protein.  
SQ SEQUENCE 44 AA; 4900 MW; CA9523A82D448591 CRC64;

Alignment Scores:  
Pred. No.: 0.00241 Length: 44  
Score: 12.00 Matches: 12  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 2.52% Indels: 0  
DB: 11 Gaps: 0

US-10-020-441-1 (1-1477) x Q99KF9 (1-44)

OY 688 CTGTCTCTGTCTGTATCTGTTGTCTGTCTCTCT 723  
|  
DB 11 LeucyValcysValserValcysValserValcys 22

RESULT 4  
O9N8S9 PRELIMINARY: PRT: 188 AA.  
ID O9N8S9  
AC O9N8S9  
DT 01-OCT-2000 (TREMBlrel. 15, Created)  
DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)  
DT 01-OCT-2000 (TREMBlrel. 15, Last annotation update)  
DE Possible (hiv-6) u1102, variant a dna, complete virion genome.  
OS Trypanosoma brucei.  
OC Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae; Trypanosoma.  
OX NCBI\_TaxID=5691;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA STRAIN=TRE0927;  
RA Hall N., Bowman S., Quail M., Ivens A.C., Kay M.P., Bray-Allen S.,  
RA Gerard N.J., Clark L.N., Harris B.R., Melville S., Lawson D.,  
RA Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.  
DR EMBL: AL359782; CAB95454.1; -  
SQ SEQUENCE 188 AA; 20899 MW; 9EAEAD143A19071 CRC64;

Alignment Scores:  
Pred. No.: 0.283 Length: 188  
Score: 10.00 Matches: 10

Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 2.10% Indels: 0  
DB: 5 Gaps: 0  
US-10-020-441-1 (1-1477) x Q9N8S9 (1-188)

OY 232 GCGTGGCGTGGCTTTGTGTGTATGC 261  
|  
DB 51 AlacysAlacysValcysValcysValcys 60

RESULT 5  
O94176 PRELIMINARY: PRT: 361 AA.  
ID O94176  
AC O94176  
DT 01-FEB-1997 (TREMBlrel. 02, Created)  
DT 01-DEC-2001 (TREMBlrel. 19, Last sequence update)  
DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)  
DE Hypothetical 40.0 kDa protein.  
GN C43G2.5.  
OS Caenorhabditis elegans.  
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;  
OC Rhabditidae; Peloderinae; Caenorhabditis.  
OX NCBI\_TaxID=6239;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=BRISTOL N2;  
RX MEDLINE=99069613; PubMed=9851916;  
RA None;  
RT "Genome sequence of the nematode C. elegans: a platform for  
RT investigating biology. The C. elegans Sequencing Consortium.";  
RL Science 282:2012-2018(1998).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=BRISTOL N2;  
RA Fulton B., Wohlmann P.;  
RT "The sequence of C. elegans cosmid C43G2.";  
RL Submitted (OCT-1996) to the EMBL/GenBank/DBJ databases.  
RN [3]  
RP SEQUENCE FROM N.A.  
RC STRAIN=BRISTOL N2;  
RA Waterston R.;  
RT "Direct Submission.";  
RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.  
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE  
CC TRYPSIN FAMILY.  
DR EMBL: U70848; AAB09110.2; -  
DR HSSP: P00766; ICHG.  
DR InterPro: IPR001314; Chymotrypsin.  
DR InterPro: IPR001254; Ser\_Protease\_Try.  
DR PRINTS: PR00722; CHYMOTRYPSIN.  
DR SMART: SM00020; TRYP\_SPC; 1.  
DR PROSITE: PS00240; TRYPSIN\_DOM; 2.  
DR PROSITE: PS00134; TRYPSIN\_HIS; UNKNOWN.1.  
DR PROSITE: PS00135; TRYPSIN\_SER; UNKNOWN.2.  
KM Hydrolyase; Hypothetical protein; Serine protease.  
SQ SEQUENCE 361 AA; 39951 MW; ECC41115BCCE7924 CRC64;

Alignment Scores:  
Pred. No.: 0.256 Length: 361  
Score: 10.00 Matches: 10  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 2.10% Indels: 0  
DB: 5 Gaps: 0  
US-10-020-441-1 (1-1477) x Q94176 (1-361)

OY 1322 GCTCCAGTGACAGTGGTGAACCTCTCTC 1351  
|  
DB 225 AlaProGlyAspSerGlyGlyProLeu 234

RESULT 6



Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 1.90% Indels: 0  
DB: 2 Gaps: 0

US-10-020-441-1 (1-1477) x 007651 (1-696)

OY 443 CTGAGTGTCTGCGATGCTCTC 417  
Db 6 LeuserValGysAlaIleGlyLeu 14

RESULT 10

O9PPG3 PRELIMINARY: PRT: 696 AA.  
ID O9PPG3  
AC 01-OCT-2000 (TREMBlrel. 15, Created)  
DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)  
DE Putative iron uptake protein.  
GN CPERA OR C10755  
OS Campylobacter jejuni.  
OC Bacteria; Proteobacteria; epsilon subdivision; Campylobacter group;  
OC Campylobacter.  
OX NCBI\_TaxID=197;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-NCTC 11168;  
RX MEDLINE=20150912; PubMed=10688204;  
RA Parkhill J., Wren B.W., Mungall K., Ketley J.M., Churcher C.,  
RA Basham D., Chillingworth T., Davies R.M., Felwell T., Holtroyd S.,  
RA Jorgels K., Karylshev A.V., Moule S., Pallen M.J., Penn C.W.,  
RA Ouzil M.A., Rajandream M.A., Rutherford K.M., van Vliet A.H.M.,  
RA Whitehead S., Barrett B.G.;  
RT "The genome sequence of the food-borne pathogen Campylobacter jejuni  
RT reveals hypervariable sequences.";  
RL Nature 403:665-668(2000).  
DR EMBL: AL139076; CAB73021.1; -  
DR InterPro: IPR000531; TonB\_boxC.  
DR Pfam: PF00593; TonB\_boxC; 1.  
KW Complete proteome.  
SQ SEQUENCE 696 AA; 77554 MW; 31E09CB15187388C CRC64;

Alignment Scores:  
Pred. No.: 2.8 Length: 696  
Score: 9.00 Matches: 9  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 1.90% Indels: 0  
DB: 16 Gaps: 0

US-10-020-441-1 (1-1477) x O9PPG3 (1-696)

OY 443 CTGAGTGTCTGCGATGCTCTC 417  
Db 6 LeuserValGysAlaIleGlyLeu 14

RESULT 11

O9TTG4 PRELIMINARY: PRT: 46 AA.  
ID O9TTG4  
AC O9TTG4;  
DT 01-MAY-2000 (TREMBlrel. 13, Created)  
DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)  
DE 01-JUN-2002 (TREMBlrel. 21, Last annotation update)  
DE Heart chimaera (Fragment).  
GN CMA.  
OS Ateles belzebuth chamek (Chamek spider monkey).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Platyrrhini; Cebidae; Ateleinae; Ateles.  
OX NCBI\_TaxID=118643;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=20169179; PubMed=10702662;  
RA Canavez F.C., Moreira M.A., Bonvicino C.R., Parham P., Seanez H.N.;

RT "Evolutionary disruptions of human syntenic groups 3, 12, 14, and 15  
RT in Ateles belzebuth chamek (Platyrrhini, Primates).";  
RT Cytogenet. Cell Genet. 87:182-188(1999).  
RL EMBL: AF099174; AAF21848.1; -  
DR HSSP: P23946; 1KLT.  
DR MEROPS: S01.140; -  
DR InterPro: IPR001254; Ser\_protease\_Try.  
DR Pfam: PF00089; trypsin.1.  
DR PROSITE: PS00135; TRYPsin\_SER; 1.  
KW Hydrolase; Serine protease.  
FT NON\_TER 1  
FT NON\_TER 1  
SQ SEQUENCE 46 AA; 5003 MW; 481EA020A265D29F CRC64;

Alignment Scores:  
Pred. No.: 51.8 Length: 46  
Score: 8.00 Matches: 8  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 1.68% Indels: 0  
DB: 6 Gaps: 0

US-10-020-441-1 (1-1477) x O9TTG4 (1-46)

OY 1328 GGTGACAGTGTGACCTCTCCTC 1351  
Db 36 GlyaspSerGlyProleu 43

RESULT 12

O9AOY9 PRELIMINARY: PRT: 105 AA.  
ID O9AOY9  
AC O9AOY9;  
DT 01-JUN-2001 (TREMBlrel. 17, Created)  
DT 01-JUN-2001 (TREMBlrel. 17, Last sequence update)  
DE 01-JUN-2001 (TREMBlrel. 17, Last annotation update)  
DE P0498A12.14 protein (OSJBA0004B13.25 protein).  
GN P0498A12.14.  
OS Oryza sativa (Rice).  
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;  
OC Ehrhartoideae; Oryzae; Oryza.  
OX NCBI\_TaxID=4530;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-CV. NIPPONBARE;  
RA Sasaki T., Matsunoto T., Yamamoto K.;  
RT "Oryza sativa nippobare(GA3) genomic DNA, chromosome 1, PAC  
RT clone: P0498A12.";  
RL Submitted (DEC-2000) to the EMBL/Genbank/DBJ databases.  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN-CV. NIPPONBARE;  
RA Sasaki T., Matsunoto T., Yamamoto K.;  
RT "Oryza sativa nippobare(GA3) genomic DNA, chromosome 1, BAC  
RT clone: OSJBA0004B13.";  
RL Submitted (DEC-2000) to the EMBL/Genbank/DBJ databases.  
DR EMBL: AP003020; BAB39985.1; -  
DR EMBL: AP003018; BAB39970.1; -  
SQ SEQUENCE 105 AA; 11120 MW; 9A9CD191A38A5A0 CRC64;

Alignment Scores:  
Pred. No.: 45.5 Length: 105  
Score: 8.00 Matches: 8  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 1.69% Indels: 0  
DB: 10 Gaps: 0

US-10-020-441-1 (1-1477) x O9AOY9 (1-105)

OY 1356 GATGGAGAGAGATCCACCTG 1333  
Db 6 AspGlyArgArgGlyProProleu 13

## RESULT 13

```

Q9UDH5 PRELIMINARY; PRT; 115 AA.
ID Q9UDH5;
AC Q9UDH5;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Chymase (Fragments).
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE.
RX MEDLINE=94194169; PubMed=8144971;
RA Schechter N.M., Wang Z.M., Blacher R.W., Lessin S.R., Lazarus G.S.,
RA Rubin H.;
RT "Determination of the primary structures of human skin chymase and
RT cathepsin G from cutaneous mast cells of urticaria pigmentosa
RT lesions."
RL J. Immunol. 152:4062-4069(1994).
DR HSP: P23946; IRL:
DR MEROPS: S01.140; -.
DR InterPro: IPR001254; Ser_protease_Try.
DR SMART: SM00020; Tryp_Spc: 1.
DR PROSITE: PS00240; TRYPsin_DOM: 1.
DR PROSITE: PS00135; TRYPsin_SER: 1.
KW Hydrolyase; Serine protease.
FT NON_TER 1 1
FT NON_CONS 35 36
FT NON_CONS 63 64
FT NON_TER 115 115
SQ SEQUENCE 115 AA; 12607 MW; 857D34B742192F77 CRC64;

Alignment Scores:
Pred. No.: 44.9 Length: 115
Score: 8.00 Matches: 8
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 1.68% Indels: 0
Gaps: 0

US-10-020-441-1 (1-1477) x Q9UDH5 (1-115)
QY 1328 GGTGACAGTGGTGACCTCTCCTC 1351
Db 68 GlyAsperGlyGlyProLeuLeu 75

RESULT 14
Q96UD9 PRELIMINARY; PRT; 120 AA.
ID Q96UD9;
AC Q96UD9;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE Hypothetical 13.5 kDa protein.
GN 15E11.155.
OS Neurospora crassa.
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
OC Sordariales; Sordariaceae; Neurospora.
OX NCBI_TaxID=5141;
RN [1]
RP SEQUENCE FROM N.A.
RA Schulte U., Algn V., Hoheisel J., Brandt P., Farthmann B., Holland R.,
RA Nyakatura G., Mewes H.W., Mannhaupt G.;
RL Submitted (DEC-2000) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RA German Neurospora genome project;
RL Submitted (NOV-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL: AL451020; CAD11362.1; -.
KW Hypothetical protein.

```

SQ SEQUENCE 120 AA; 13456 MW; 5F562457492E6DE2 CRC64;

## Alignment Scores:

Pred. No.: 44.6 Length: 120  
 Score: 8.00 Matches: 8  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 1.69% Indels: 0  
 DB: 3 Gaps: 0

US-10-020-441-1 (1-1477) x Q96UD9 (1-120)

QY 175 CTCTTACCGCATCAGTGGCATG 152

Db 58 LeuLeuProGluSerLeuAlaMet 65

## RESULT 15

```

Q9EQ13 PRELIMINARY; PRT; 128 AA.
ID Q9EQ13;
AC Q9EQ13;
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE Mast cell protease-9 (Fragment).
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6;
RX MEDLINE=20580915; PubMed=11141502;
RA Ge Y., Jippo T., Lee Y.-M., Adachi S., Kitamura Y.;
RT "Independent influence of strain difference and ml transcription
RT factor on the expression of mouse mast cell chymases."
RL Am. J. Pathol. 158:281-292(2001).
DR EMBL: AF302235; AAC44997.1; -.
DR HSP: P00770; 3RP2.
DR MEROPS: S01.304; -.
DR InterPro: IPR001254; Ser_protease_Try.
DR Pfam: PF00089; trypsin: 1.
DR SMART: SM00020; Tryp_Spc: 1.
DR PROSITE: PS00240; TRYPsin_DOM: 1.
DR PROSITE: PS00135; TRYPsin_SER: 1.
KW Hydrolyase; Protease; Serine protease.
FT NON_TER 1 1
FT NON_TER 128 128
SQ SEQUENCE 128 AA; 13497 MW; F516645FBFA38F67 CRC64;

Alignment Scores:
Pred. No.: 44.2 Length: 128
Score: 8.00 Matches: 8
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 1.68% Indels: 0
Gaps: 0

US-10-020-441-1 (1-1477) x Q9EQ13 (1-128)
QY 1328 GGTGACAGTGGTGACCTCTCCTC 1351
Db 82 GlyAsperGlyGlyProLeuLeu 89

RESULT 16
Q9YW35 PRELIMINARY; PRT; 132 AA.
ID Q9YW35;
AC Q9YW35;
DT 01-MAY-1999 (TrEMBLrel. 10, Created)
DT 01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE ORF MSV057 hypothetical protein.
GN MSV057.
OS Melanoplus sanguinipes entomopoxvirus (MsePV).

```

OC Viruses; dsDNA viruses, no RNA stage; Poxviridae; Entomopoxvirinae;  
 OC Entomopoxvirus B.  
 OX NCBI\_TaxID=83191;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-TUCSON;  
 RX MEDLINE=99102612; PubMed=9847359;  
 RA Alfonso C.L., Tulman E.R., Lu Z., Oma E., Kutish G.F., Rock D.L.;  
 RT "The genome of *Melanoplus sanguinipes* entomopoxvirus."; *J. Virol.* 73:533-552(1999).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-TUCSON;  
 RA Alfonso C.L., Tulman E.R., Lu Z., Oma E., Kutish G.F., Rock D.L.;  
 RL Submitted (MAY-1998) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: AF063866; AAC97821.1;  
 KW Hypothetical protein.  
 SQ SEQUENCE 132 AA; 13990 MW; CB5487A53549C4E2 CRC64;

Alignment Scores:  
 Pred. No.: 43.9 Length: 132  
 Score: 8.00 Matches: 8  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 1.68% Indels: 0  
 DB: 12 Gaps: 0

US-10-020-441-1 (1-1477) x Q9YV35 (1-132)

OY 669 CAGATGTCAGTCAGTTATCTGTG 692  
 Db 16 GlnmetserValserTyrSerVal 23

RESULT 17  
 Q92GJ4 PRELIMINARY; PRT; 156 AA.  
 ID Q92GJ4  
 AC Q92GJ4  
 DT 01-DEC-2001 (TREMblrel. 19, Created)  
 DT 01-DEC-2001 (TREMblrel. 19, Last sequence update)  
 DT 01-DEC-2001 (TREMblrel. 19, Last annotation update)  
 DE Hypothetical protein RC1129.  
 GN RC1129.  
 OS Rickettsia conorii.  
 OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;  
 OC Rickettsiaceae; Rickettsiae; Rickettsia.  
 OX NCBI\_TaxID=781;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-MALISH 7;  
 RX MEDLINE=21442074; PubMed=11557893;  
 RA Ogata H., Audic S., Renesto-Audiffren P., Fournier P.-E., Barbe V.,  
 RA Sanson D., Roux V., Cossart P., Weissbach J., Claverie J.-M.,  
 RA Raoult D.;  
 RT "Mechanisms of evolution in *Rickettsia conorii* and *R. prowazekii*."; *Science* 293:2093-2098(2001).  
 RL EMBL: AE008662; AAL03667.1;  
 KW Hypothetical protein; Complete proteome.  
 SQ SEQUENCE 156 AA; 18208 MW; 8E0612BDFD4ED5B9 CRC64;

Alignment Scores:  
 Pred. No.: 42.8 Length: 156  
 Score: 8.00 Matches: 8  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 1.68% Indels: 0  
 DB: 16 Gaps: 0

US-10-020-441-1 (1-1477) x Q92GJ4 (1-156)

OY 1451 TTGATTTTGACGCTCCAAATATT 1474  
 Db 62 LeuAspHeValArgSerAsnIle 69

RESULT 18  
 Q8VDE7 PRELIMINARY; PRT; 176 AA.  
 ID Q8VDE7  
 AC Q8VDE7  
 DT 01-MAR-2002 (TREMblrel. 20, Created)  
 DT 01-MAR-2002 (TREMblrel. 20, Last sequence update)  
 DT 01-JUN-2002 (TREMblrel. 21, Last annotation update)  
 DE Chymase 2.  
 OS *Mus musculus* (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.  
 OX NCBI\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Chu W., Johnson D.A., Musich P.R.;  
 RT "Molecular cloning and characterization of mouse mast cell chymases."; *Submitted (JAN-1992) to the EMBL/GenBank/DBJ databases.*  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RA Chu W.;  
 RT "Molecular cloning and characterization of mouse mast cell chymases."; *Thesis (1991), Unknown Institution.*  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=92287966; PubMed=1376147;  
 RA Chu W., Johnson D.A., Musich P.R.;  
 RT Biochim. Biophys. Acta 1121:83-87(1992).  
 DR EMBL: M68899; AAA39493.1;  
 DR InterPro: IPR001314; Chymotrypsin.  
 DR InterPro: IPR001254; Ser-protease-Try.  
 DR Pfam: PF00089; trypsin.1.  
 DR PRINTS: PR00722; CHYMOTRYPSIN.  
 DR SMART: SM00020; TRYP-Sec.1.  
 DR PROSITE: PS50240; TRYP-SER\_DOM.1.  
 DR PROSITE: PS00135; TRYP-SER; UNKNOWN.1.  
 SQ SEQUENCE 176 AA; 19173 MW; DCF586204DD18475 CRC64;

Alignment Scores:  
 Pred. No.: 42 Length: 176  
 Score: 8.00 Matches: 8  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 1.68% Indels: 0  
 DB: 11 Gaps: 0

US-10-020-441-1 (1-1477) x Q8VDE7 (1-176)

OY 1328 GGTGACAGTGTGACGCTCTCTC 1351  
 Db 130 GlyAspSerGlyGlyProLeuLeu 137

RESULT 19  
 Q9XYW2 PRELIMINARY; PRT; 214 AA.  
 ID Q9XYW2  
 AC Q9XYW2  
 DT 01-NOV-1999 (TREMblrel. 12, Created)  
 DT 01-NOV-1999 (TREMblrel. 12, Last sequence update)  
 DT 01-MAR-2002 (TREMblrel. 20, Last annotation update)  
 DE Enterokinase-like protein (Fragment).  
 GN EKL.  
 OS *Schistosoma mansoni* (Blood fluke).  
 OC Eukaryota; Metazoa; Platyhelminthes; Trematoda; Digenea; Strigeidida;  
 OC Schistosomatidae; Schistosomatidae; Schistosoma.  
 OX NCBI\_TaxID=6183;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-PUERTO RICAN;  
 RA El-Bassiouni N.E., Bailey J.A., Blanton R.E.;  
 RT "Developmental expression of serine protease activity in *Schistosoma mansoni*."; *Submitted (FEB-1999) to the EMBL/GenBank/DBJ databases.*  
 RL Submitted (FEB-1999) to the EMBL/GenBank/DBJ databases.  
 CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE

```
CC TRYPsin FAMILY.
DR EMBL: AF128677; AAD32612.1; -.
DR HSSP: P00763; IDPO.
DR InterPro: IPR001314; Chymotrypsin.
DR InterPro: IPR001254; Ser.protease_Try.
DR Pfam: PF00089; trypsin_1.
DR PRINTS: PR00722; CHYMOTRYPsin.
DR SMART: SM00020; TRYp_Spc; 1.
DR PROSITE: PS50240; TRYPsin_DOM; 1.
DR PROSITE: PS00135; TRYPsin_SER; 1.
KW Hydrolyase; Kinase; Serine protease.
FT NON_TER 1
SQ SEQUENCE 214 AA; 24334 MW; 20B3E8F2987F30C3 CRC64;

Alignment Scores:
Pred. No.: 40.8 Length: 214
Score: 8.00 Matches: 8
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 1.08% Indels: 0
DB: Gaps: 0

US-10-020-441-1 (1-1477) x Q9XYW2 (1-214)
QY 1325 CCAGGTGACAGTGTGACCTCTC 1348
Db 144 ProGlyAspserGlyGlyProLeu 151

RESULT 20
Q9XYW2 PRELIMINARY; PRT; 223 AA.
ID Q9XYW2;
AC Q9XYW2;
DT 01-NOV-1999 (TrEMBLrel. 12, Created)
DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DE 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Chymotrypsin-like serine protease (Fragment).
OS SP-4.
GN Ctenocephalides felis (Cat flea).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Siphonaptera; Pulicidae;
OC Fleecineae; Ctenocephalides.
OX NCBI_TaxID=7515;
RN [1]
RP SEQUENCE FROM N.A.
RA Gaines P.J., Sampson C.M., Rushlow K.E., Stiegler G.L.;
RT "Cloning of a family of serine protease genes from the cat flea
RT Ctenocephalides felis.";
RL Insect Mol. Biol. 81:11-22(1999).
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE
CC TRYPsin FAMILY.
DR EMBL: AF053916; AAD21836.1; -.
DR HSSP: P00766; ICHG.
DR InterPro: IPR001314; Chymotrypsin.
DR InterPro: IPR001254; Ser.protease_Try.
DR Pfam: PF00089; trypsin_1.
DR PRINTS: PR00722; CHYMOTRYPsin.
DR SMART: SM00020; TRYp_Spc; 1.
DR PROSITE: PS50240; TRYPsin_DOM; 1.
DR PROSITE: PS00135; TRYPsin_SER; 1.
KW Hydrolyase; Kinase; Serine protease.
FT NON_TER 1
SQ SEQUENCE 223 AA; 24844 MW; C34EBE9455DD7DE9 CRC64;

Alignment Scores:
Pred. No.: 40.5 Length: 223
Score: 8.00 Matches: 8
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 1.08% Indels: 0
DB: Gaps: 0

US-10-020-441-1 (1-1477) x Q9XYW2 (1-223)

QY 1328 GGTGACAGTGTGACCTCTCTC 1351
Db 177 GlyAspserGlyGlyProLeu 184

RESULT 21
Q9YBY4 PRELIMINARY; PRT; 223 AA.
ID Q9YBY4;
AC Q9YBY4;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DE 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE CG11836 protein (LD25630P).
GN CG11836.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachyera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BERKELEY.
RX MEDLINE=20196006; Pubmed=10731132;
RA Adams M.D., Celnik S.E., Holt R.A., Evans C.A., Cocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazer R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfankoch C., Baldwin D.,
RA Baller R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borikova D., Botchan M.R., Bouck J., Brokstein P., Brotler P.,
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablo B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferriera S., Fleischmann W.,
RA Foster C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodde A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ideyama C.,
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Sytkas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wasserman D.A., Weinstein G.M., Weissenbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yen R.-F., Zaveri J.S., Zhan W., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of Drosophila melanogaster.";
RL Science 287:2165-2195(2000).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=BERKELEY.
RA Stapleton M., Brokstein P., Hong L., Agbayani A., Carlson J.,
RA Champe M., Chavez C., Dorsett V., Farfan D., Frise E., George R.,
RA Gonzalez M., Guarin H., Li P., Liao G., Miranda A., Mungall C.J.,
RA Nuno J., Pacleb J., Paragas V., Park S., Phouanavong S., Wan K.,
RA Yu C., Lewis S.E., Rubin G.M., Celnik S.,
RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE
CC TRYPsin FAMILY.
DR EMBL: AF003750; AAF56393.1; -.

```

DR EMBL: AY061333; AAL28881.1; .  
DR HSSP: P00763; IDPO.  
DR FlyBase: FBgn0039272; CG11836.  
DR InterPro: IPR001314; Chymotrypsin.  
DR InterPro: IPR001254; Ser\_protease\_Try.  
DR Pfam: PF00089; trypsin\_1.  
DR PRINTS: PR00722; CHYMOTRYPSIN.  
DR SMART: SM00020; TRYP\_SPE; 1.  
DR PROSITE: PS50240; TRYPsin\_DOM; 1.  
DR PROSITE: PS00134; TRYPsin\_HIS; 1.  
DR PROSITE: PS00135; TRYPsin\_SER; 1.  
DR HydroLase: Serine protease.  
SQ SEQUENCE 223 AA; 24625 MW; 1AE47014802B1DF5 CRC64;  
  
Alignment Scores:  
Pred. No.: 40.5 Length: 223  
Score: 8.00 Matches: 8  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 1.68% Indels: 0  
DB: 5 Gaps: 0  
  
US-10-020-441-1 (1-1477) x Q9VBY4 (1-223)  
OY 1328 GGTGACAGTGTGACCTCTCCTC 1351  
Db 167 GLYSPserGlyGlyProLeuLeu 174  
|||||  
  
RESULT 22  
Q9TXE6 PRELIMINARY; PRT; 232 AA.  
AC Q9TXE6;  
DT 01-MAY-2000 (TREMBLrel. 13, Created)  
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)  
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)  
DE ALKALIPHILIC serine protease P-TRIC-TRYPsin-like protease.  
OS Bombyx mori (Silk moth).  
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
OC Pterygota; Neoptera; Endopterygota; Lepidoptera; Glossata; Ditrysia;  
OC Bombycoidea; Bombycidae; Bombyx.  
OX NCBI\_TaxID=7091;  
RN [1]  
RP SEQUENCE.  
RX MEDLINE=93215816; PubMed=8462672;  
RA Sasaki T., Hishida T., Ichikawa K., Asari S.;  
RT "Amino acid sequence of alkaliphilic serine protease from silkworm,  
RT Bombyx mori, larval digestive juice.";  
RL FEBS Lett. 320:35-37(1993).  
CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE  
CC TRYPsin FAMILY.  
DR HSSP: P00763; IDPO.  
DR MEROPS: S01.112; .  
DR InterPro: IPR001314; Chymotrypsin.  
DR InterPro: IPR001254; Ser\_protease\_Try.  
DR Pfam: PF00089; trypsin\_1.  
DR PRINTS: PR00722; CHYMOTRYPSIN.  
DR SMART: SM00020; TRYP\_SPE; 1.  
DR PROSITE: PS50240; TRYPsin\_DOM; 1.  
DR PROSITE: PS00134; TRYPsin\_HIS; UNKNOWN\_1.  
DR PROSITE: PS00135; TRYPsin\_SER; 1.  
KM HydroLase: Serine protease.  
SQ SEQUENCE 232 AA; 25413 MW; 4B4A54A90E2434C3 CRC64;  
  
Alignment Scores:  
Pred. No.: 40.3 Length: 232  
Score: 8.00 Matches: 8  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 1.68% Indels: 0  
DB: 5 Gaps: 0  
  
US-10-020-441-1 (1-1477) x Q9TXE6 (1-232)

OY 1328 GGTGACAGTGTGACCTCTCCTC 1351  
Db 167 GLYSPserGlyGlyProLeuLeu 194  
|||||  
  
RESULT 23  
Q9GTX7 PRELIMINARY; PRT; 235 AA.  
AC Q9GTX7;  
DT 01-MAR-2001 (TREMBLrel. 16, Created)  
DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)  
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)  
DE late trypsin (Fragment).  
OS Aedes albopictus (Forest day mosquito).  
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
OC Pterygota; Neoptera; Endopterygota; Diptera; Nematocera; Culicoidae;  
OC Aedes.  
OX NCBI\_TaxID=7160;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Chao X., Tongyan Z., Baolin L.;  
RT "Cloning and sequencing of cDNA coding for late trypsin induced by  
RT blood meal in Aedes albopictus."  
RL submitted (MAY-2000) to the EMBL/Genbank/DBJ databases.  
CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE  
CC TRYPsin FAMILY.  
DR EMBL: AF268665; AAG10005.1; .  
DR HSSP: P00763; IDPO.  
DR MEROPS: S01.112; .  
DR InterPro: IPR001314; Chymotrypsin.  
DR InterPro: IPR001254; Ser\_protease\_Try.  
DR Pfam: PF00089; trypsin\_1.  
DR PRINTS: PR00722; CHYMOTRYPSIN.  
DR SMART: SM00020; TRYP\_SPE; 1.  
DR PROSITE: PS50240; TRYPsin\_DOM; 1.  
DR PROSITE: PS00134; TRYPsin\_HIS; UNKNOWN\_1.  
DR PROSITE: PS00135; TRYPsin\_SER; 1.  
KW HydroLase: Serine protease.  
FT NON\_TER 1  
SQ SEQUENCE 235 AA; 25232 MW; AB39A28C264A0604 CRC64;  
  
Alignment Scores:  
Pred. No.: 40.2 Length: 235  
Score: 8.00 Matches: 8  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 1.68% Indels: 0  
DB: 5 Gaps: 0  
  
US-10-020-441-1 (1-1477) x Q9GTX7 (1-235)  
OY 121 GTACTGACAGCTGTCATTGTGTT 144  
Db 42 ValLeuThrAlaGlyHisCysVal 49  
|||||  
  
RESULT 24  
Q9TYH3 PRELIMINARY; PRT; 236 AA.  
AC Q9TYH3;  
DT 01-MAY-2000 (TREMBLrel. 13, Created)  
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)  
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)  
DE Serine protease SmSP1 light chain (Fragment).  
OS Schistosoma mansoni (Blood fluke).  
OC Eukaryota; Metazoa; Platyhelminthes; Trematoda; Digenea; Strigeiida;  
OC Schistosomatidae; Schistosomatidae; Schistosoma.  
OX NCBI\_TaxID=6183;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA STRAIN=PORTO R1C0.  
RT "Identification of a developmentally regulated schistosoma mansoni  
RT serine protease homologous to mouse plasma kallikrein and human Factor  
RT I.";

RL Submitted (SEP-1998) to the EMBL/GenBank/DBJ databases.  
CC -I- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE  
TRYPSIN FAMILY.  
DR EMBL: AJ011562; CAA09692.1; -.  
DR HSSP: P00766; 1CHG.  
DR InterPro: IPR001314; Chymotrypsin.  
DR InterPro: IPR001254; Ser-protease\_try.  
DR Pfam: PF00089; trypsin; 1.  
DR PRINTS: PR00722; CHYMOTRYPsin.  
DR SMART: SM00020; TRYP-SPC; 1.  
DR PROSITE: PS50240; TRYPSIN\_DOM; 1.  
DR PROSITE: PS00134; TRYPSIN\_HIS; UNKNOWN\_1.  
DR PROSITE: PS00135; TRYPSIN\_SER; 1.  
KW Hydrolyase; Protease; Serine protease.  
FT NON\_TER  
SQ SEQUENCE 236 AA: 26224 MW: FF9C3CDFBAC2BA4 CRC64;  
  
Alignment Scores:  
Pred. No.: 40.1 Length: 236  
Score: 8.00 Matches: 8  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 1.68% Indels: 0  
DB: 5 Gaps: 0  
  
US-10-020-441-1 (1-1477) x Q9TYH3 (1-236)  
QY 1328 GGTGACAGTGTGACCTCTCTC 1351  
DB 187 GLYASPERGLYGLYProleuLeu 194  
|||||  
RESULT 25  
Q9EQT2 PRELIMINARY; PRT; 239 AA.  
AC Q9EQT2;  
DT 01-MAR-2001 (TREMBLrel. 16, Created)  
DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)  
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)  
DE Myonase.  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
OX NCBI\_TaxID=10090;  
RN (1)  
RP SEQUENCE FROM N.A.  
RC TISSUE-SKELETAL MUSCLE;  
RA Hori S.;  
RL Submitted (DEC-2000) to the EMBL/GenBank/DBJ databases.  
RN (2)  
RP SEQUENCE FROM N.A.  
RC TISSUE-SKELETAL MUSCLE;  
RA MEDLINE-98207054; PubMed-9538257;  
RA Hori S., Ohtani S., Hori C., Nohikara K.;  
RT "Purification and characterization of myonase from x-chromosome linked  
muscular dystrophic mouse skeletal muscle.";  
RL J. Biochem. 123:650-658(1998).  
CC -I- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE  
TRYPSIN FAMILY.  
CC EMBL: AB051900; BAB18732.1; -.  
DR HSSP: P00770; 3RP2.  
DR MEROPS: S01.149; -.  
DR InterPro: IPR001314; Chymotrypsin.  
DR InterPro: IPR001254; Ser-protease\_try.  
DR Pfam: PF00089; trypsin; 1.  
DR PRINTS: PR00722; CHYMOTRYPsin.  
DR SMART: SM00020; TRYP-SPC; 1.  
DR PROSITE: PS50240; TRYPSIN\_DOM; 1.  
DR PROSITE: PS00134; TRYPSIN\_HIS; UNKNOWN\_1.  
DR PROSITE: PS00135; TRYPSIN\_SER; 1.  
KW Hydrolyase; Serine protease.  
SQ SEQUENCE 239 AA: 26379 MW: 8539FC6971F9F049 CRC64;  
  
Alignment Scores:

Pred. No.: 40.1 Length: 239  
Score: 8.00 Matches: 8  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 1.68% Indels: 0  
DB: 11 Gaps: 0

US-10-020-441-1 (1-1477) x Q9EQT2 (1-239)

QY 1328 GGTGACAGTGTGACCTCTCTC 1351  
DB 193 GLYASPERGLYGLYProleuLeu 200  
|||||

RESULT 26

Q8WL74 PRELIMINARY; PRT; 242 AA.  
AC Q8WL74;  
DT 01-MAR-2002 (TREMBLrel. 20, Created)  
DT 01-MAR-2002 (TREMBLrel. 20, Last sequence update)  
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)  
DE NADH dehydrogenase subunit F (Fragment).  
GN NDHF.  
OS Mionandra camareoides.  
OC Chloroplast.  
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;  
OC eurosids I; Malpighiales; Malpighiaceae; Mionandra.  
OX NCBI\_TaxID=151861;  
RN (1)  
RP SEQUENCE FROM N.A.  
RA Davis C.C., Anderson W.R., Donoghue M.J.;  
RT "Phylogeny of Malpighiaceae: evidence from chloroplast ndhF and trnL-F  
nucleotide sequences.";  
RT Am. J. Bot. 88:1830-1846(2001).  
DR EMBL: AF351080; AAL35094.1; -.  
DR InterPro: IPR002128; Oxidored\_q1\_C.  
DR Pfam: PF01010; oxidored\_q1\_C; 1.  
KW Chloroplast.  
FT NON\_TER  
SQ SEQUENCE 242 AA: 28185 MW: D6BD0CAD4CA90DAC CRC64;

Alignment Scores:

Pred. No.: 40 Length: 242  
Score: 8.00 Matches: 8  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 1.69% Indels: 0  
DB: 8 Gaps: 0

US-10-020-441-1 (1-1477) x Q8WL74 (1-242)

QY 570 ATACCAACATGAACACCACT 547  
DB 62 IlerhrasnasngluasnrInSer 69  
|||||

RESULT 27

Q9EPR0 PRELIMINARY; PRT; 246 AA.  
AC Q9EPR0;  
DT 01-MAR-2001 (TREMBLrel. 16, Created)  
DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)  
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)  
DE Mast cell protease-9.  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
OX NCBI\_TaxID=10090;  
RN (1)  
RP SEQUENCE FROM N.A.  
RC STRAIN=C57BL/6;  
RA MEDLINE-20580915; PubMed-11141502;  
RA Ge Y., Jippo T., Lee Y.-M., Adachi S., Kitamura Y.;



"Independent influence of strain difference and mi transcription factor on the expression of mouse mast cell chymases.",  
 RL Am. J. Pathol. 158:281-292(2001).  
 CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE  
 TRYPsin FAMILY.  
 DR EMBL: AY007568; AAG24502.1; -.  
 DR HSSP: P00770; 3RP2.  
 DR MEROPS: S01.304; -.  
 DR InterPro: IPR001314; Chymotrypsin.  
 DR InterPro: IPR001254; Ser-protease\_Try.  
 DR Pfam: PF00089; trypsin; 1.  
 DR PRINTS: PR00722; CHYMOTRYPsin.  
 DR SMART: SM00020; Tryp\_Spc; 1.  
 DR PROSITE: PS50240; TRYPsin\_DOM; 1.  
 DR PROSITE: PS00134; TRYPsin\_HIS; UNKNOWN\_1.  
 DR PROSITE: PS00135; TRYPsin\_SER; 1.  
 DR Hydrolase: Protease; Serine protease.  
 SQ SEQUENCE 246 AA; 26636 MW; 346F0B6FA05CBEC5 CRC64;

## Alignment Scores:

Pred. No.:	39.9	Length:	246
Score:	8.00	Matches:	8
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	1.68%	Indels:	0
DB:	11	Gaps:	0

US-10-020-441-1 (1-1477) x Q9EP09 (1-246)

OY 1328 GGTGACAGTGTGACCTCTCCTC 1351  
 |||||||  
 DB 200 GlyaspserGlyGlyProleuLeu 207

## RESULT 28

Q9EP09 PRELIMINARY; PRT; 246 AA.  
 AC Q9EP09;  
 DT 01-MAR-2001 (TREMBLrel. 16, Created)  
 DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)  
 DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)  
 DE Mouse mast cell protease-4.  
 OS Mus musculus (Mouse).  
 CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 CC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OX NCBI\_TaxId=10090.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=C57BL/6;  
 RX MEDLINE=20580915; PubMed=11141502;  
 RA Ge Y., Jippo T., Lee Y.-M., Adachi S., Kitamura Y.;  
 RT "Independent influence of strain difference and ml transcription  
 factor on the expression of mouse mast cell chymases.";  
 RL Am. J. Pathol. 158:281-292(2001).  
 CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE  
 TRYPsin FAMILY.  
 DR EMBL: AY007569; AAG24503.1; -.  
 DR HSSP: P00770; 3RP2.  
 DR MEROPS: S01.149; -.  
 DR InterPro: IPR001314; Chymotrypsin.  
 DR InterPro: IPR001254; Ser-protease\_Try.  
 DR Pfam: PF00089; trypsin; 1.  
 DR PRINTS: PR00722; CHYMOTRYPsin.  
 DR SMART: SM00020; Tryp\_Spc; 1.  
 DR PROSITE: PS50240; TRYPsin\_DOM; 1.  
 DR PROSITE: PS00134; TRYPsin\_HIS; UNKNOWN\_1.  
 DR PROSITE: PS00135; TRYPsin\_SER; 1.  
 DR Hydrolase: Protease; Serine protease.  
 SQ SEQUENCE 246 AA; 27196 MW; 8DC6DDDB667E8257 CRC64;

## Alignment Scores:

Pred. No.:	39.9	Length:	246
Score:	8.00	Matches:	8
Percent Similarity:	100.00%	Conservative:	0

Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	1.68%	Indels:	0
DB:	11	Gaps:	0

US-10-020-441-1 (1-1477) x Q9EP09 (1-246)

OY 1328 GGTGACAGTGTGACCTCTCCTC 1351  
 |||||||  
 DB 200 GlyaspserGlyGlyProleuLeu 207

## RESULT 29

Q9R2C8 PRELIMINARY; PRT; 246 AA.  
 AC Q9R2C8;  
 DT 01-MAY-2000 (TREMBLrel. 13, Created)  
 DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)  
 DT 01-MAR-2002 (TREMBLrel. 20, Last annotation update)  
 DE Mast cell protease 5 precursor (Fragment).  
 GN MCP-5.  
 OS Rattus norvegicus (Rat).  
 CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 CC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
 OX NCBI\_TaxId=10116;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=SPRAGUE DAWLEY;  
 RX MEDLINE=97149430; PubMed=8996238;  
 RA Lutzeltschhab C., Pejler G., Aveskog M., Hellman L.;  
 RT "Secretory granule proteases in rat mast cells. Cloning of 10  
 RT different serine proteases and a carboxypeptidase A from various rat  
 RT mast cell populations.";  
 RL J. Exp. Med. 185:13-29(1997).  
 CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE  
 TRYPsin FAMILY.  
 DR EMBL: U67908; AAB48261.1; -.  
 DR HSSP: P23946; 1KLT.  
 DR MEROPS: S01.150; -.  
 DR InterPro: IPR001314; Chymotrypsin.  
 DR InterPro: IPR001254; Ser-protease\_Try.  
 DR Pfam: PF00089; trypsin; 1.  
 DR PRINTS: PR00722; CHYMOTRYPsin.  
 DR SMART: SM00020; Tryp\_Spc; 1.  
 DR PROSITE: PS50240; TRYPsin\_DOM; 1.  
 DR PROSITE: PS00134; TRYPsin\_HIS; UNKNOWN\_1.  
 DR PROSITE: PS00135; TRYPsin\_SER; 1.  
 DR Hydrolase: Protease; Serine protease; Signal.  
 FT NON-TER 1  
 FT SIGNAL 1  
 FT CHAIN <1 18 POTENTIAL.  
 FT CHAIN 19 246 MAST CELL PROTEASE 5.  
 SQ SEQUENCE 246 AA; 27438 MW; A525FC41BDFEF87 CRC64;

## Alignment Scores:

Pred. No.:	39.9	Length:	246
Score:	8.00	Matches:	8
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	1.68%	Indels:	0
DB:	11	Gaps:	0

US-10-020-441-1 (1-1477) x Q9R2C8 (1-246)

OY 1328 GGTGACAGTGTGACCTCTCCTC 1351  
 |||||||  
 DB 200 GlyaspserGlyGlyProleuLeu 207

## RESULT 30

Q91VB1 PRELIMINARY; PRT; 246 AA.  
 AC Q91VB1;  
 DT 01-DEC-2001 (TREMBLrel. 19, Created)  
 DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)  
 DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)  
 DE Mast cell protease 10.

```

GN MOP10.
OS Mus musculus (mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-BALB/C, AND BALB/CJ; TISSUE=JEJUNUM;
RA Hunt J.E., McNeil P., Austen F., Gurish M., Stevens R.L.;
RT "Molecular cloning of the cDNA and gene encoding mouse mast cell
RT protease 10."
RL Submitted (MAR-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF361940; AAK51075.1; -.
DR EMBL: AF361939; AAK51074.1; -.
DR MEROPS; S01.304; -.
DR InterPro: IPR001254; Ser-protease TRY.
DR Pfam: PF00089; trypsin.1.
DR PROSITE: PS50240; TRYPSIN_DOM; 1.
DR PROSITE: PS00134; TRYPSIN_HIS; UNKNOWN_1.
DR PROSITE: PS00135; TRYPSIN_SER; UNKNOWN_1.
KW Hydrolase; Protease; Serine protease.
SQ SEQUENCE 246 AA; 26857 MW; 3E2005E0E809FFD3 CRC64;

Alignment Scores:
Pred. No.: 39.9 Length: 246
Score: 8.00 Matches: 8
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 1.68% Indels: 0
DB: 11 Gaps: 0

US-10-020-441-1 (1-1477) x Q91VB1 (1-246)
QY 1328 GGTGACAGTGTGACCTCTCTC 1351
Db 200 GLYASPERGLYGLYPROLEULEN 207

RESULT 31
Q17039 PRELIMINARY; PRT; 247 AA.
ID Q17039;
AC Q17039;
DT 01-NOV-1996 (TREMBLrel. 01, Created)
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE Serine proteinase (Fragment).
OS Anopheles gambiae (African malaria mosquito).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Nematocera; Culicoidae;
OC Anopheles.
OX NCBI_TaxID=7165;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-G3;
RA MEDLINE=96209489; PubMed=8630536;
RX Siden-Klamos I., Skavdis G., Rubio J., Papaianakakis G., Louis C.;
RT "Isolation and characterisation of three serine protease genes in the
RT mosquito Anopheles gambiae."
RL Insect Mol. Biol. 5:61-71(1996).
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE
CC TRYPSIN FAMILY.
DR EMBL: Z49813; CA69967.1; -.
DR HSSP: P00763; IDPO.
DR InterPro: IPR001314; Chymotrypsin.
DR InterPro: IPR001254; Ser-protease TRY.
DR Pfam: PF00089; trypsin.1.
DR PRINTS: PR00089; trypsin.1.
DR PRINTS: PR00839; V8PROTEASE.
DR SMART: SM00020; TRYP_SPC; 1.
DR PROSITE: PS50240; TRYPSIN_DOM; 1.
DR PROSITE: PS00134; TRYPSIN_HIS; UNKNOWN_1.
DR PROSITE: PS00135; TRYPSIN_SER; 1.
KW Hydrolase; Serine protease.

```

```

FT NON_TER 1 1 SERINE PROTEINASE.
SQ CHAIN 10 247
SQ SEQUENCE 247 AA; 27210 MW; 50417BF2CFB98D9 CRC64;

Alignment Scores:
Pred. No.: 39.9 Length: 247
Score: 8.00 Matches: 8
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 1.68% Indels: 0
DB: 5 Gaps: 0

US-10-020-441-1 (1-1477) x Q17039 (1-247)
QY 1328 GGTGACAGTGTGACCTCTCTC 1351
Db 190 GLYASPERGLYGLYPROLEULEN 197

RESULT 32
Q8WL73 PRELIMINARY; PRT; 247 AA.
ID Q8WL73
AC Q8WL73;
DT 01-MAR-2002 (TREMBLrel. 20, Created)
DT 01-MAR-2002 (TREMBLrel. 20, Last sequence update)
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE NADH dehydrogenase subunit F (Fragment).
OS NDHF.
OC Cordobia argentea.
OC Cordobia argentea.
OC Chloroplast.
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids I; Malpighiales; Malpighiaceae; Cordobia.
OX NCBI_TaxID=151807;
RN [1]
RP SEQUENCE FROM N.A.
RA Davis C.C., Anderson W.R., Donoghue M.J.;
RT "Phylogeny of Malpighiaceae: evidence from chloroplast ndhF and trnL-F
RT nucleotide sequences."
RL Am. J. Bot. 88:1830-1846(2001).
DR EMBL: AF351081; AAL35095.1; -.
DR InterPro: IPR002128; Oxidored_q1_C.
DR Pfam: PR01010; oxidored_q1_C; 1.
KW Chloroplast.
FT NON_TER 1 1
FT NON_TER 247 247
SQ SEQUENCE 247 AA; 28631 MW; 513F2A3D169FF0DE CRC64;

Alignment Scores:
Pred. No.: 39.9 Length: 247
Score: 8.00 Matches: 8
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 1.69% Indels: 0
DB: 8 Gaps: 0

US-10-020-441-1 (1-1477) x Q8WL73 (1-247)
QY 570 ATAACAACATGAACACACAGT 547
Db 65 IlerhrsnasnngluasnThrSer 72

RESULT 33
Q08732 PRELIMINARY; PRT; 247 AA.
ID Q08732
AC Q08732;
DT 01-JUL-1997 (TREMBLrel. 04, Created)
DT 01-JUL-1997 (TREMBLrel. 04, Last sequence update)
DT 01-MAR-2002 (TREMBLrel. 20, Last annotation update)
DE Chymase 1.
OS Mesocricetus auratus (Golden hamster).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Cricetinae;
OC Mesocricetus.

```

```

OX NCBI_TaxID=10036;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE-97354983; PubMed-9211178;
RA Shota N., Fukamizu A., Takai S., Okunishi H., Murakami K.,
  Miyazaki M.;
RT "Activation of angiotensin II-forming chymase in the cardiomyopathic
  hamster heart.";
RL J. Hypertens. 15:431-440(1997).
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE
  TRYPSIN FAMILY.
DR EMBL: D8517; BAA1932.1; -.
DR HSSP: P00770; 3RP2.
DR MEROPS: S01.149; -.
DR InterPro: IPR001314; Chymotrypsin.
DR InterPro: IPR001254; Ser_Protease_Try.
DR Pfam: PF00089; Trypsin_1.
DR PRINTS: PR00722; CHYMOTRYPSIN.
DR SMART: SM00020; TRYP_SPE; 1.
DR PROSITE: PS50240; TRYPSIN_DOM; 1.
DR PROSITE: PS00134; TRYPSIN_HIS; UNKNOWN_1.
DR PROSITE: PS00135; TRYPSIN_SER; 1.
KW Hydrolase; Serine protease.
FT CHAIN 21 247
SQ SEQUENCE 247 AA; 26926 MW; 75E8EFB1E0F4C7E CRC64;

Alignment Scores:
Pred. No.: 39.9 Length: 247
Score: 8.00 Matches: 8
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 1.68% Indels: 0
DB: 11 Gaps: 0

US-10-020-441-1 (1-1477) x 008732 (1-247)
QY 1328 GGTGACAGTGTGACCTCTCCTC 1351
DB 200 GlyaspserGlyGlyProleuLeu 207

RESULT 34
070164
ID 070164 PRELIMINARY; PRT; 247 AA.
AC 070164;
DT 01-AUG-1998 (TREMBLrel. 07, Created)
DT 01-AUG-1998 (TREMBLrel. 07, Last sequence update)
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE Chymase 2.
OS Mesocricetus auratus (Golden hamster).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Cricetinae;
OC Mesocricetus.
OX NCBI_TaxID=10036;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE-98324904; PubMed-9657983;
RA Shota N., Fukamizu A., Okunishi H., Takai S., Murakami K.,
  Miyazaki M.;
RT "Cloning of the gene and cDNA for hamster chymase 2, and expression of
  chymase 1, chymase 2 and angiotensin-converting enzyme in the terminal
  stage of cardiomyopathic hearts.";
RL Biochem. J. 333:417-424(1998).
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE
  TRYPSIN FAMILY.
DR EMBL: AB007622; BAA28615.1; -.
DR HSSP: P23946; 1KLT.
DR MEROPS: S01.150; -.
DR InterPro: IPR001314; Chymotrypsin.
DR InterPro: IPR001254; Ser_Protease_Try.
DR Pfam: PF00089; Trypsin_1.
DR PRINTS: PR00722; CHYMOTRYPSIN.
DR SMART: SM00020; TRYP_SPE; 1.
DR PROSITE: PS50240; TRYPSIN_DOM; 1.
DR PROSITE: PS00134; TRYPSIN_HIS; UNKNOWN_1.
DR PROSITE: PS00135; TRYPSIN_SER; 1.
KW Hydrolase; Protease; Serine protease.
FT CHAIN 247 AA; 27105 MW; BB89DB314383D88 CRC64;

SQ SEQUENCE 247 AA; 27105 MW; BB89DB314383D88 CRC64;

Alignment Scores:
Pred. No.: 39.9 Length: 247
Score: 8.00 Matches: 8
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 1.68% Indels: 0
DB: 11 Gaps: 0

US-10-020-441-1 (1-1477) x 035342 (1-247)
QY 1328 GGTGACAGTGTGACCTCTCCTC 1351
DB 200 GlyaspserGlyGlyProleuLeu 207

RESULT 36
070500

```

```

DR PROSITE: PS00134; TRYPSIN_HIS; UNKNOWN_1.
DR PROSITE: PS00135; TRYPSIN_SER; 1.
KW Hydrolase; Serine protease.
FT CHAIN 22 247
SQ SEQUENCE 247 AA; 27510 MW; 0C371AE44CD91540 CRC64;

Alignment Scores:
Pred. No.: 39.9 Length: 247
Score: 8.00 Matches: 8
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 1.68% Indels: 0
DB: 11 Gaps: 0

US-10-020-441-1 (1-1477) x 070164 (1-247)
QY 1328 GGTGACAGTGTGACCTCTCCTC 1351
DB 201 GlyaspserGlyGlyProleuLeu 208

RESULT 35
035342
ID 035342 PRELIMINARY; PRT; 247 AA.
AC 035342;
DT 01-JAN-1998 (TREMBLrel. 05, Created)
DT 01-JAN-1998 (TREMBLrel. 20, Last sequence update)
DT 01-MAR-2002 (TREMBLrel. 20, Last annotation update)
DE Skeletal muscle neutral protease mekralin.
OS Mesocricetus auratus (Golden hamster).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Cricetinae;
OC Mesocricetus.
OX NCBI_TaxID=10036;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-BIO 53:58;
RA Margossian S.S., Umeda P.K., Sciaky D., Caulfield J.B., Melendez J.A.,
  Martino L., Chantler P.D., Slayter H.S.;
RT "Myopathic hamster mRNA for the skeletal muscle neutral protease
  mekralin.";
RL Mol. Cell. Biochem. 0:0-0(1997).
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE
  TRYPSIN FAMILY.
DR EMBL: AF020767; AAB91992.1; -.
DR HSSP: P00770; 3RP2.
DR MEROPS: S01.255; -.
DR InterPro: IPR001314; Chymotrypsin.
DR InterPro: IPR001254; Ser_Protease_Try.
DR Pfam: PF00089; Trypsin_1.
DR PRINTS: PR00722; CHYMOTRYPSIN.
DR SMART: SM00020; TRYP_SPE; 1.
DR PROSITE: PS50240; TRYPSIN_DOM; 1.
DR PROSITE: PS00134; TRYPSIN_HIS; UNKNOWN_1.
DR PROSITE: PS00135; TRYPSIN_SER; 1.
KW Hydrolase; Protease; Serine protease.
FT CHAIN 247 AA; 27105 MW; BB89DB314383D88 CRC64;

SQ SEQUENCE 247 AA; 27105 MW; BB89DB314383D88 CRC64;

Alignment Scores:
Pred. No.: 39.9 Length: 247
Score: 8.00 Matches: 8
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 1.68% Indels: 0
DB: 11 Gaps: 0

US-10-020-441-1 (1-1477) x 035342 (1-247)
QY 1328 GGTGACAGTGTGACCTCTCCTC 1351
DB 200 GlyaspserGlyGlyProleuLeu 207

RESULT 36
070500

```

```

ID 070500 PRELIMINARY; PRT; 247 AA.
AC 070500;
DT 01-AUG-1998 (TREMBLrel. 07, Created)
DT 01-AUG-1998 (TREMBLrel. 07, Last sequence update)
DT 01-MAR-2002 (TREMBLrel. 20, Last annotation update)
DE Vascular chymase.
VCH.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=SPRAGUE-DAWLEY; TISSUE=PULMONARY ARTERY;
RX MEDLINE=21153933; PubMed=11254670;
RA Guo C., Ju H., Leung D., Massaeli H., Shi M., Rabinovitch M.;
RT "A novel vascular smooth muscle chymase is upregulated in hypertensive
RT rats.";
RL J. Clin. Invest. 107:703-715(2001).
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE
CC TRYPSIN FAMILY.
CC EMBL: AF063851; AAC16657.1; -.
DR HSSP: P00770; 3RP2.
DR MEROPS: S01.095; -.
DR InterPro: IPR001314; Chymotrypsin.
DR InterPro: IPR001254; Ser_protease_Try.
DR Pfam: PF00089; trypsin_1.
DR PRINTS: PR00722; CHYMOTRYPSIN.
DR SMART: SM00020; TRYP_SPC: 1.
DR PROSITE: PS50240; TRYPsin_DOM: 1.
DR PROSITE: PS00134; TRYPsin_HIS; UNKNOWN_1.
DR PROSITE: PS00135; TRYPsin_SER: 1.
KW Hydrolyase; Serine protease.
SQ SEQUENCE 247 AA; 27278 MW; FF913C1103D6D9F9 CRC64;

Alignment Scores:
Pred. No.: 39.9 Length: 247
Score: 8.00 Matches: 8
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 1.68% Indels: 0
DB: 11 Gaps: 0

US-10-020-441-1 (1-1477) x 070500 (1-247)
QY 1328 GGTGACAGTGGTGACCTCTCTC 1351
    |||||||
DB 200 GLYAsperglyglyProleuLeu 207

RESULT 37
O34289 PRELIMINARY; PRT; 255 AA.
ID 034289;
AC 034289;
DT 01-JAN-1998 (TREMBLrel. 05, Created)
DT 01-JAN-1998 (TREMBLrel. 05, Last sequence update)
DT 01-MAR-2002 (TREMBLrel. 20, Last annotation update)
DE Serine protease-like protein precursor.
OS Salvelinus fontinalis (Brook trout) (Brook char).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Euteleostei;
OC Protacanthopterygii; Salmoniformes; Salmonidae; Salvelinus.
OX NCBI_TaxID=8038;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=PREOVULATORY OVARIAN TISSUE;
RX MEDLINE=98206794; PubMed=9546717;
RA Hajnik C.A., Goetz F.W., Hsu S.Y., Sokal N.;
RT "Characterization of a ribonucleic acid transcript from the brook
RT trout (Salvelinus fontinalis) ovary with structural similarities to
RT mammalian adipisin/complement factor D and tissue kallikrein, and the
RT effects of kallikrein-like serine proteases on follicle contraction.";
RL Biol. Reprod. 58:887-897(1998).
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE

```

```

CC TRYPSIN FAMILY.
DR EMBL: AF005026; AAC17927.1; -.
DR HSSP: P00761; 1EPT.
DR MEROPS: S01.191; -.
DR InterPro: IPR001314; Chymotrypsin.
DR InterPro: IPR001254; Ser_protease_Try.
DR Pfam: PF00089; trypsin_1.
DR PRINTS: PR00722; CHYMOTRYPSIN.
DR SMART: SM00020; TRYP_SPC: 1.
DR PROSITE: PS50240; TRYPsin_DOM: 1.
DR PROSITE: PS00134; TRYPsin_HIS; UNKNOWN_1.
DR PROSITE: PS00135; TRYPsin_SER: 1.
KW Hydrolyase; Protease; Serine protease; Signal.
FT CHAIN 1
FT SIGNAL 1
FT CHAIN 36
FT SIGNAL 22
SQ SEQUENCE 255 AA; 27658 MW; 9FB8A0B2C3E648FA CRC64;

Alignment Scores:
Pred. No.: 39.7 Length: 255
Score: 8.00 Matches: 8
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 1.68% Indels: 0
DB: 13 Gaps: 0

US-10-020-441-1 (1-1477) x 034289 (1-255)
QY 1328 GGTGACAGTGGTGACCTCTCTC 1351
    |||||||
DB 207 GLYAsperglyglyProleuLeu 214

RESULT 38
O54426 PRELIMINARY; PRT; 264 AA.
ID 054426;
AC 054426;
DT 01-NOV-1996 (TREMBLrel. 01, Created)
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE Rubredoxin reductase (Fragment).
GN ALKA.
OS Xanthomonas maltophilia (Pseudomonas maltophilia) (Stenotrophomonas
OS maltophilia).
OC Bacteria; Proteobacteria; gamma subdivision; Xanthomonas group.
OC Stenotrophomonas.
OX NCBI_TaxID=40324;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=N246;
RX MEDLINE=96136269; PubMed=8573125;
RA Lee N.R., Hwang M.O., Jung G.H., Kim Y.S., Min K.H.;
RT "Physical structure and expression of alkA encoding alkane
RT hydroxylase and rubredoxin reductase from Pseudomonas maltophilia.";
RL Biochem. Biophys. Res. Commun. 218:17-21(1996).
DR EMBL: U40234; AAA97867.1; -.
FT NON_TER 264
FT SEQUENCE 264 AA; 29726 MW; 315C2E7AC872CAB CRC64;

Alignment Scores:
Pred. No.: 39.5 Length: 264
Score: 8.00 Matches: 8
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 1.69% Indels: 0
DB: 2 Gaps: 0

US-10-020-441-1 (1-1477) x 054426 (1-264)
QY 124 GTAAGCTCTGCTGAGACTAGTG 101
    |||||||
DB 102 ValLeuLeuSerLeuArgLeuVal 109

RESULT 39
O9VAX6

```

ID O9VAX6 PRELIMINARY: PRT: 265 AA.  
AC O9VAX6;  
DT 01-MAY-2000 (TREMBLrel. 13, Created)  
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)  
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)  
DE CG4815 protein.  
GN CG4815.  
OS Drosophila melanogaster (fruit fly).  
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
OC Ephydroidea; Drosophilidae; Drosophila.  
OX NCBI\_Taxid=7227;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=BERKELEY;  
RX MEDLINE=20196006; PubMed=10731132;  
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,  
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galie R.F.,  
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,  
RA Sutton G.G., Mortman J.R., Yandell M.D., Zhang Q., Chen L.X.,  
RA Brandon R.C., Rogers Y.-H.C., Blazey R.G., Chape M., Pfeiffer B.D.,  
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,  
RA Abdl J.F., Agbayan A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,  
RA Baller R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,  
RA Beeson K.Y., Benos P.V., Bernan B.P., Bhandari D., Bolshakov S.,  
RA Borkova D., Botchan M.R., Bouck J., Brockstein P., Brotlier P.,  
RA Butts K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,  
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,  
RA de Pablo S., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,  
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,  
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,  
RA Foster C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glasser K.,  
RA Glodok A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,  
RA Harris N.L., Harvey D., Helman T.J., Hernandez J.R., Honck J.,  
RA Hostin L.D., Houston K.A., Howland T.J., Wei M.-H., Iqbalwag C.,  
RA Jaisl M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,  
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,  
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,  
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,  
RA Mekulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,  
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,  
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Paclebb J.M.,  
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,  
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,  
RA Shue B.C., Siden-Klamos I., Simpson M., Skupski M.P., Smith T.,  
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,  
RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,  
RA Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissenbach J.,  
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,  
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,  
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,  
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;  
RT "The genome sequence of Drosophila melanogaster."  
RL Science 287:2185-2195(2000).  
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE  
CC TRYPsin FAMILY.  
CC EMBL: AEO03763; AAF56771.1; -  
DR HSP: P00761; IAKS;  
DR FLYBase: FBgn0039368; CG4815.  
DR InterPro: IPR001314; Chymotrypsin.  
DR InterPro: IPR001254; Ser.protease\_Try.  
DR Pfam: PF00089; trypsin.1.  
DR PRINTS: PR00722; CHYMOTRYPSIN.  
DR SMART: SM00020; TRYP-Spc.1.  
DR PROSITE: PS50240; TRYPsin\_DOM.1.  
DR PROSITE: PS00134; TRYPsin\_HIS.1.  
KW Hydroxylase; Serine protease.  
SO SEQUENCE 265 AA; 30080 MW; 10A04171B2E6A397 CRC64;

Alignment Scores:  
Pred. No.: 39.4 Length: 265  
Score: 8.00 Matches: 8  
Percent Similarity: 100.00% Conservative: 0

Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 1.68% Indels: 0  
DB: 5 Gaps: 0  
US-10-020-441-1 (1-1477) x O9VAX6 (1-265)

OY 1328 GGTGACAGTGGTGGACCTCTCCTC 1351  
DB 215 GLYASPSERGILYProleulin 222

## RESULT 40

ID O17030 PRELIMINARY: PRT: 268 AA.  
AC O17030;  
DT 01-NOV-1996 (TREMBLrel. 01, Created)  
DT 01-FEB-1997 (TREMBLrel. 02, Last sequence update)  
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)  
DE Serine protease.  
OS Anopheles gambiae (African malaria mosquito).  
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
OC Pterygota; Neoptera; Endopterygota; Diptera; Nematocera; Culicoidae;  
OC Anophelinae.  
OX NCBI\_Taxid=7165;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=G3; TISSUE=MIDGUT;  
RX MEDLINE=97075119; PubMed=8917545;  
RA Dimopoulos G.M., Richman A., della Torre A., Kafatos F.C., Louis C.,  
RT "Identification and characterization of differentially expressed cDNAs  
RT of the vector mosquito, Anopheles gambiae."  
CC Proc. Natl. Acad. Sci. U.S.A. 93:13066-13071(1996).  
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE  
CC TRYPsin FAMILY.  
DR EMBL: Z69978; CAA93818.1; -  
DR HSP: P00763; IDPO.  
DR InterPro: IPR001314; Chymotrypsin.  
DR InterPro: IPR001254; Ser.protease\_Try.  
DR Pfam: PF00089; trypsin.1.  
DR PRINTS: PR00722; CHYMOTRYPSIN.  
DR SMART: SM00020; TRYP-Spc.1.  
DR PROSITE: PS50240; TRYPsin\_DOM.1.  
DR PROSITE: PS00134; TRYPsin\_HIS.1.  
DR PROSITE: PS00135; TRYPsin\_SER.1.  
KW Hydroxylase; Protease; Serine protease.  
SO SEQUENCE 268 AA; 29176 MW; 7BBE8462EF1PD8BE CRC64;

Alignment Scores:  
Pred. No.: 39.4 Length: 268  
Score: 8.00 Matches: 8  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 1.68% Indels: 0  
DB: 5 Gaps: 0

US-10-020-441-1 (1-1477) x O17030 (1-268)

OY 121 GTACTCAGACGCTGCTATGCTT 144  
DB 69 VALLTEINRAlaGlyHiscysVal 76

Search completed: April 10, 2003, 07:56:32  
Job time : 174 secs

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.4 p5.4578  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - protein search, using frame\_plus\_n2p model

Run on: April 9, 2003, 13:16:18 / Search time 87 seconds

(without alignments)  
4524.395 Million cell updates/sec

Title: US-10-020-441-1

Perfect score: 2686  
Sequence: 1 gtgtcaaccgtgtgacg.....ctgtacgtccacatattga 1477

Scoring table:

BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 1816940

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Command line parameters:  
-MODE=frame+ n2p model -DEV=xlp  
-Q=/cgn2\_1/USPTO.epool/US10020441/runat 09042003 121210 1263/app query.fasta\_1.1671  
-DB=GeneSeq\_101002 -QFMT=faasn -SUFFIX=rag -MINMATCH=0.1 -DOFCL=0  
-LOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=blomum2 -TRANS=human40.cdi  
-LIST=45 -DOCALIGN=200 -THR SCORE=DCT -THR MAX=100 -THR MIN=0 -ALIGN=40  
-MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MITLEN=0 -MAXLEN=2000000000  
-USR=US10020441 @CGN 1.1 114 @runat 09042003 121210 1263 -NCPU=6 -ICPU=3  
-NO\_XLPPY -NO\_MMIP -LARGEQUERY -NEG\_SCORES=0 -WAIT -LONGLOG -DEV TIMEOUT=120  
-MARK TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7  
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :  
A\_GeneSeq\_101002:\*  
1: /SIDS2/gcgdata/geneSeq/geneSeq-emb1/AA1980.DAT.\*  
2: /SIDS2/gcgdata/geneSeq/geneSeq-emb1/AA1981.DAT.\*  
3: /SIDS2/gcgdata/geneSeq/geneSeq-emb1/AA1982.DAT.\*  
4: /SIDS2/gcgdata/geneSeq/geneSeq-emb1/AA1983.DAT.\*  
5: /SIDS2/gcgdata/geneSeq/geneSeq-emb1/AA1984.DAT.\*  
6: /SIDS2/gcgdata/geneSeq/geneSeq-emb1/AA1985.DAT.\*  
7: /SIDS2/gcgdata/geneSeq/geneSeq-emb1/AA1986.DAT.\*  
8: /SIDS2/gcgdata/geneSeq/geneSeq-emb1/AA1987.DAT.\*  
9: /SIDS2/gcgdata/geneSeq/geneSeq-emb1/AA1988.DAT.\*  
10: /SIDS2/gcgdata/geneSeq/geneSeq-emb1/AA1989.DAT.\*  
11: /SIDS2/gcgdata/geneSeq/geneSeq-emb1/AA1990.DAT.\*  
12: /SIDS2/gcgdata/geneSeq/geneSeq-emb1/AA1991.DAT.\*  
13: /SIDS2/gcgdata/geneSeq/geneSeq-emb1/AA1992.DAT.\*  
14: /SIDS2/gcgdata/geneSeq/geneSeq-emb1/AA1993.DAT.\*  
15: /SIDS2/gcgdata/geneSeq/geneSeq-emb1/AA1994.DAT.\*  
16: /SIDS2/gcgdata/geneSeq/geneSeq-emb1/AA1995.DAT.\*  
17: /SIDS2/gcgdata/geneSeq/geneSeq-emb1/AA1996.DAT.\*  
18: /SIDS2/gcgdata/geneSeq/geneSeq-emb1/AA1997.DAT.\*  
19: /SIDS2/gcgdata/geneSeq/geneSeq-emb1/AA1998.DAT.\*  
20: /SIDS2/gcgdata/geneSeq/geneSeq-emb1/AA1999.DAT.\*  
21: /SIDS2/gcgdata/geneSeq/geneSeq-emb1/AA2000.DAT.\*  
22: /SIDS2/gcgdata/geneSeq/geneSeq-emb1/AA2001.DAT.\*  
23: /SIDS2/gcgdata/geneSeq/geneSeq-emb1/AA2002.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	552	20.6	106	21	AAV67587
2	152	5.7	5532	21	AAAB3752
3	152	5.7	5532	21	AAAG5267
4	138	5.2	6239	21	AAAB3750
5	138	5.2	6239	21	AAAG5265
6	136	5.1	307	22	AAAB5124
7	133	5.0	227	22	ABG14399
8	131	4.9	598	22	ABG23595
9	130.5	4.9	190	20	AAAG68539
10	129.5	4.9	375	22	ABG14053
11	128.5	4.8	3972	21	AAAB3749
12	128.5	4.8	3972	22	AAAG5264
13	128.5	4.8	3972	22	AAAG5268
14	126.5	4.7	190	20	AAAG68540
15	126.5	4.7	1795	22	ABAB69806
16	126.5	4.7	1939	22	ABG20505
17	124.5	4.7	1171	22	ABAB64387
18	123.5	4.6	247	23	ABAB7898
19	123	4.6	245	23	AAU12005
20	123	4.6	245	23	AAU12023
21	123	4.6	245	23	AAU12024
22	123	4.6	249	23	AAU12006
23	123	4.6	249	23	AAU12019
24	123	4.6	249	23	AAU12020
25	123	4.6	273	19	AAAG4238
26	123	4.6	273	19	AAAG63174
27	123	4.6	691	23	AAAB14348
28	121	4.5	957	21	AAAG59288
29	121	4.5	957	21	AAAG4513
30	120.5	4.5	733	19	AAAG9430
31	119	4.4	275	19	AAAG4237
32	119	4.4	275	19	AAAG63173
33	118.5	4.4	245	22	ABG28019
34	118	4.4	270	8	ABP70759
35	118	4.4	270	23	AAAB7691
36	117.5	4.4	895	22	ABG15781
37	117	4.4	244	20	AAAG5925
38	117	4.4	245	23	AAU12008
39	117	4.4	245	23	AAU12015
40	117	4.4	245	23	AAU12016
41	117	4.4	249	21	AAAG5011
42	117	4.4	249	23	AAU12007
43	117	4.4	249	23	AAU12011
44	117	4.4	249	23	AAU12012
45	117	4.4	267	19	AAAG4241

## ALIGNMENTS

RESULT 1  
AAV67587  
ID AAV67587 standard; Protein: 106 AA.  
XX  
AC AAV67587;  
XX  
DT 19-JUN-2000 (first entry)  
XX  
DE S. mansoni cercarial elastase protein fragment.  
XX  
KW Schistosoma parasite; fusion protein; cercarial elastase; vaccine;  
KW antibody response; schistosomicide.  
XX  
OS Schistosoma mansoni.  
XX  
EP992582-A2.  
XX  
PD 12-APR-2000.  
XX

PF 05-OCT-1999; 99EP-0307832.  
XX  
PR 07-OCT-1998; 98GB-0021821.  
XX  
PA (UYWA-) UNIV WALES BANGOR.  
XX  
PI Doenhoff M, Sayers J;  
XX  
DR WPI: 2000-259136/23.  
DR N-PSDB; AAZ90608.  
XX  
PT New vaccine for treatment of Schistosoma infections contains a  
PT recombinant fusion protein comprising cercarial elastase fused  
PT to bacterial, phage or viral protein -  
XX  
PS Claim 3; Page 17; 26pp; English.  
XX  
CC The invention provides a vaccine comprising a recombinant fusion protein  
CC capable of eliciting immunity against Schistosoma parasites. The fusion  
CC protein comprises the 27 or 28 kDa cercarial elastase sequence of  
CC S. mansoni or an active fragment, homologue or variant, fused to a  
CC bacterial, phage or viral protein. The vaccine containing the fusion  
CC protein has been found to induce a significantly increased antibody  
CC response against schistosoma infections, compared to the use of S.  
CC mansoni cercarial elastase in its native form. The present sequence  
CC represents the S. mansoni cercarial elastase protein fragment encoded by  
CC exon 2 of the cercarial elastase gene, used in the fusion protein of the  
CC invention.  
XX  
SQ Sequence 106 AA;

Alignment Scores:  
Pred. No.: 2.4e-44 Length: 106  
Score: 552.00 Matches: 106  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 20.55% Indels: 0  
DB: 21 Gaps: 0

US-10-020-441-1 (1-1477) x AAY67587 (1-106)

QY 310 GTTATTCTTCTACACTGAGATGGCGACCAAGGATCCATCCCAACCGTCTGA 369  
Db 1 ValSerPheLeuThrLeuArgAsnGlyAspGlnGlnGlyLeuHisGlnProSerGly 20  
QY 370 GTTAGGGGCGACAGATACATGCCCTCTGTATGTGGCGACAGAGAGACCAATC 429  
Db 21 ValLysValAlaProGlyTyrMetProSerCysMetSerAlaArgGlnArgProIle 40  
QY 430 GCACAGACACTGATGATTCGATATTCGATATTCGATATTCGATATTCGATATTC 489  
Db 41 AlaGlnThrLeuSerGlyPheAspIleAlaIleValMetLeuAlaGlnMetAlaAsnLeu 60  
QY 490 CAGAGTGAATCAGAGTGAATCAGTGTGCGACAGCCATCGGATATCCCGACCTGGA 549  
Db 61 GlnSerGlyIleArgValIleSerLeuProGlnProSerAspIleProProGlyThr 80  
QY 550 GGTGTTTCATTTGGTGTATGGAAGGATGATACGACCGGATCCGTCAGTAAAGAT 609  
Db 81 GlyValPheIleValGlyTyrGlyArgAspAspAspAspProSerArgLysAsn 100  
QY 610 GGTGAATATTTGAAGAA 627  
Db 101 GlyGlyIleLeuLys 106  
RESULT 2  
AAB23752  
ID AAB23752 standard; Protein; 5532 AA.  
XX  
AC AAB23752;  
XX  
DT 10-JAN-2001 (first entry)  
XX

DE S. avermitilis avermectin aglycon synthase protein SEQ ID NO:6.  
XX  
KW Streptomyces avermitilis; avermectin aglycon synthase; biosynthesis;  
KW multifunctional enzyme; polyketide; avermectin; veterinary drug;  
KW agrochemical.  
XX  
OS Streptomyces avermitilis.  
XX  
PN MO200050605-A1.  
XX  
PD 31-AUG-2000.  
XX  
PF 23-FEB-2000; 2000WO-JP01041.  
XX  
PR 24-FEB-1999; 99JP-0046961.  
XX  
PA (KITA ) KITASATO INST.  
XX  
PI Omura S, Ikeda H;  
XX  
DR WPI: 2000-565458/52.  
DR N-PSDB; AAA92302.  
XX  
CC The present sequence represents an avermectin aglycon synthase protein.  
CC Also described are: (1) polypeptides encoded by all or part of the DNA;  
CC (2) expression vectors containing the DNA; (3) host cells transformed by  
CC the vectors; (4) preparation of the polypeptides by culture of the  
CC transformants; (5) preparation of avermectin aglycon or its derivatives  
CC by culture of transformed avermectin-producing microorganisms; and (6)  
CC oligonucleotides of 5-60 bases in length containing sense or antisense  
CC sequences from the avermectin aglycon synthase DNA. The enzymes are  
CC useful for the production of modified forms of avermectin and of the  
CC intermediates in its biosynthesis, for use as drugs, veterinary drugs  
CC and agrochemicals.  
XX  
SQ Sequence 5532 AA;

Alignment Scores:  
Pred. No.: 2.54e-05 Length: 5532  
Score: 152.00 Matches: 103  
Percent Similarity: 31.58% Conservative: 41  
Best Local Similarity: 22.59% Mismatches: 130  
Query Match: 5.70% Indels: 182  
DB: 21 Gaps: 20

US-10-020-441-1 (1-1477) x AAB23752 (1-5532)

QY 1358 GGGATGGAGGAGAGGTCCACCACTGACCGGGTAACTGTCCGAATTTCTGAC 1299  
Db 2436 GlyMetAlaHisGlyLeuTyrHisThrIleProValPheAlaIleAsn---Asp 2454  
QY 1298 CTGCTTACACATATAGATGTCATTCGTCATTCGTCATTCATTAATGCGCTC 1239  
Db 2455 IleCysThrHisLeu---AspProHisLeuAspHisProLeuLeuProLeuLeuThrGln 2473  
QY 1238 GACCTGTGTGTCAAACCAACCAACCAACCAACCAACCAACCAACCAACCAAC 1179  
Db 2474 AspPro-----AsnThrGlnAspHisThrThrLeuGlnGlnAlaIleAla--- 2488  
QY 1178 ATGAGAAAGAACAAACCAACCAACCAACCAACCAACCAACCAACCAACCAAC 1119  
Db 2488 ----- 2488  
QY 1118 CAATGGGCAATATATCAACCTTTACTACTACCCACACAACATAGTCCACATACAC 1059  
Db 2489 ---LeuLeuGlnGlnThrProTyrAlaGlnPro----- 2498



OY	1058	ATGTCACATAATTCAGCAAGATTCAGTCCAT	-----	1023
Db	2459	-----AAlaLeuPheAlaPheGlnValAlaLeu	HisArgLeuLeuThrAspGlyGlnHis	2516
OY	1023	-----	-----	1023
Db	2517	ILethrProHisGlyGlyAlaGlyHisSerLeuGlyGly	LeuThrAlaAlaHisLeuAla	2536
OY	1022	-----	-----	996
Db	2537	GIYLeuLeuThrLeuThrAspAlaThrThrLeuLeu	LeuThrGlnArgAlaThrLeuMetGln	2556
OY	995	ACACATTCAGCCCGCTACTTCATGATACACAA	CACACTTCGATCGTTGATACATATCTCA	936
Db	2557	ThimetProProGlyThrMetThrThrLeuHis	ThrThr-----	2563
OY	935	TGAGCTCTCTCCCTCCCAAGGCTGCAATCAT	CCCATCCATCCATCCATCCATCC-----	885
Db	2570	-----	-----	2586
OY	884	-----	-----	858
Db	2585	AlaIleAlaAlaIleAsnThrProThrSerLeuVal	IlleSerGlyThrProHisThrVal	2604
OY	857	TCACATTCAGTCACCC-----	-----	843
Db	2605	GlnHisIleThrThrLeuGlyGlnGlnGlnGly	LeuThrThrLeuProThrAsn	2624
OY	842	-----	-----	795
Db	2625	HisAlaPheHisSerProHisThrAsnProIle	LeuAsnGlnLeuHisGlnHisThrGln	2644
OY	794	CTGTACAAATGCTCCCAACACACAGTTGTT	ATTATCAAGGCTCTGCACAAATCAACAA	735
Db	2645	ThrLeuThrThrGlnHisProProHisHisThr	ProLeuLeuIleThrAla-----	2662
OY	734	CGATTCAG-----	-----	717
Db	2662	AspArgLeuLeuLeuThrProHisGlyThrThr	GlnGlnAlaArgAsnThrValAspAla	2682
OY	716	GACACACAAACAG-----	-----	660
Db	2682	AlThrThrThrGlnThrGlnThrLeuHisGln	HisValGly-----	2694
OY	659	TGAGTCACTGGTATTATTCACAACTCACT	CTCTTCACAAATATTCACAT-----	608
Db	2694	rThrThrTleuLeuLeuGlyProAsnAsnThr	ThrLeuThrHisHisAsnLeuThr	2714
OY	607	-----	-----	571
Db	2714	AsnThrProThrThrThrThrLeuThrHis	ThrHisProHisHisHisLeu-----	2734
OY	570	--ATACACAAATGAAGAAACACCAAGTTC	CAAGGCGGAGATATCCGATCGCTGGCAGA	513
Db	2734	euleThrHisLeuAlaLeuThrThrThrThr	ThrThrHisProHisHisGly-----	2751
OY	512	CTGATCACTGGATTCACACTGTGAATGACA	CTTTTGTGAGC-----	471
Db	2751	HisHisHisHisAsnGlnProHisHisHis	HisLeuAsnLeuProThrThrGlyProPhe	2771
OY	470	AGCATTCACAAATTCGAATATGCAATCC	ACTGATTC-----	426
Db	2771	HisHisHisHisGlyThrLeuGlyLeuSerThr	GlnProGlyAlaGlyAsnValSerHisAla	2790
OY	425	GCTCTCTCTCTCGTCCGACATACAAAGAG	GGCATGATCTGGTGCCACCTTAATCTCA	366
Db	2791	GlyLeu-----	-----	2806
OY	365	GACGGTGGTGGATGAGGCTGTT-----	GGTGCGCATTCG	329
Db	2806	HisGlnGlyHisArgGlyLeuLeuThrHis	GlyAlaThrLeuGlnLeu-----	2819

ID	AA065267		
AC	AA065267	standard; Protein: 5532 AA.	
XX			
XX	AA065267;		
XX			
DT	04-DEC-2001	(first entry)	
XX			
DE	Streptomyces avermiltillis	protein SEQ ID NO: 7.	
XX			
KM	Avermectin aglycone synthase; AAS; avermectin derivative;		
KM	drug production; veterinary drug; pesticide.		
XX			
OS	Streptomyces avermiltillis.		
XX			
PN	MO200162939-A1.		
PD			
XX	30-AUG-2001.		
XX			
PF	23-FEB-2001; 2001MO-JP01381.		
XX			
PR	24-FEB-2000; 2000JP-0047405.		
XX			
PA	(KYOW ) KYOWA HAKKO KOGYO KK.		
XX	(KITA ) KITASATO INST.		
PI	Endo H, Yamaguchi H, Kanda Y, Hashimoto S, Omura S, Ikeda H;		
XX			
DR	WPI; 2001-582053/65.		
XX	N-PSDB; AAH79278.		
PT	New modified avermectin aglycone synthase derived from Streptomyces		
XX	avermiltillis used in production of 22,23-dihydroavermectin B1a used in		
XX	drugs and pesticides -		
PS	Claim 4; Page 217-235; 257pp; Japanese.		
XX			
CC	The present invention relates to the production of modified derivatives		
CC	of avermectin aglycone synthase (AAS) derived from Streptomyces		
CC	avermiltillis. The activity of an acyl carrier protein (ACP),		
CC	beta-ketacyl carrier protein synthase (KS), acyltransferase (AT),		
CC	beta-ketacyl carrier protein reductase (KR), dehydratase (DH), enoyl		
CC	reductase (ER) and/or thioesterase (TE) domain may be reduced or		
CC	suppressed. The process can be used in the production of drugs, veterinary		
CC	drugs and pesticides. The present sequence is an S. avermiltillis protein		
XX			
SO	Sequence 5532 AA;		
	Alignment Scores:		
	Pred. No.:	2,54e-05	Length: 5532
	Score:	152.00	Matches: 103
	Percent Similarity:	31.58%	Conservative: 41
	Best Local Similarity:	22.59%	Mismatches: 130
	Query Match:	5.70%	Indels: 182
	Query Match:	22	Gaps: 20
US-10-020-441-1	(1-1477) x	AA065267	(1-5532)
OY	1358	GGATGGGAGGAGGAGGATCCACCACTGTCACCTGGAGCGGTAATGTCGCAAAATTCTGAC	1299
DB	2436	glymetalanhisgylleutyrlhsthrhlsprovalpheaiaalaaleuasn--asp	2454
OY	1298	CTGCTTCACACATATGATGATTCGTCATTGGTCGATGTCGCAATTCATATGATGCGTC	1239
DB	2455	ileeythrhrslsu---aspprohisleuaslhrhisproleuasnproleuleuthngln	2473
OY	1238	GACCCCTGTGTGTCGCAACCAACCAACCAACAGTACAGTACAGTACAGTACAG	1179
DB	2474	asppro---asnthnglnasppthrthrleugluhualaaiaaiaa---asp	2488
OY	1178	ATGAGAAAGAAACAAACGAAACGAAACGAAACGAAACGAAACGAAACGAAATAT	1119
DB	2488	-----asp	2488

QY	1118	CAAGTGGGCAATATTCACACCCCTTTACAGTACCCGACAAACAAATATGACCAATATGACATAC	1059
Db	2489	---LeuLeuGlnGlnThrProGlyAlaGlnPro-----	2498
QY	1058	ATGTGCACATATATTCGACGATTCACAGTCACTTCCAT-----	1023
Db	2499	-----AlaLeuPheAlaPheGlnValAlaLeuHisArgLeuLeuThrAspGlyGlyHis	2526
QY	1023	-----	1023
Db	2517	IleThrProHisGlyGlyAlaGlyHisSerLeuGlyGlyIleThrAlaHisLeuAla	2536
QY	1022	-----CTGACACCTGTCGAAACATATGACA	996
Db	2537	GlyIleLeuThrIleThrAspAlaThrThrLeuIleIleThrGlnArgAlaThrLeuMetGln	2556
QY	995	ACACATTCGACCCCTTACTTCATGATGACACAAACATCCATCCGTTTGACATAGATATTTCA	936
Db	2557	ThrMetProProGlyThrIleMetThrThrLeuHisIleThr-----	2569
QY	935	TGAGTCTTCTCTCCGCAAGTGGTCTCATGATCCATGATCCATGATCC-----	885
Db	2570	-----ProHisIleIleThrHisHisIleThrAlaHisGlyHisAspLeu	2584
QY	884	-----ATCCCATCCAAATCTACCTGACATCC	858
Db	2585	AlaIleAlaAlaIleAsnThrProThrSerLeuValIleSerGlyThrProHisThrVal	2604
QY	857	TCACATGCATCCATCC-----	843
Db	2605	GlnHisIleThrThrIleLeuGlyGlnGlnGlnGlyIleLeuThrIleYsThrLeuProThrAsn	2624
QY	842	-----TCACCTCAACCAACATGACCCGATGGAATGACCTCTGCTACTGAA	795
Db	2625	HisAlaPheHisSerProHisIleThrAsnProIleLeuAsnGlnIleHisIleGlnHisThrGln	2644
QY	794	CTGTGCACAAATCTCCAAACACAGTTGTTATATGAGGCTGTGACCAATGACAAAC	735
Db	2645	ThrIleThrGlyHisProProHisIleThrProLeuIleThrAla-----AsnThr-ProPr	2662
QY	734	GGATCAG-----GTAGACAGACA	717
Db	2662	AspSerIleLeuLeuThrProHisGlyTrpThrGlnGlnAlaArgAsnThrValAspIleAla	2682
QY	716	GACACACAAACAG--ATACACAGACACACAGAAATACTGATGACATCTGACATGACAC	660
Db	2682	athThrThrGlnThrGlnThrLeuHisGlnHisGly-----ValThr	2694
QY	659	TGACGACATGCTGTTATATTCACAAACATCACTTTCTCAATATTCACAAAT-----	608
Db	2694	rThrGlyIleGlyLeuGlyProAspAsnThrIleThrIleThrThrHisHisAsnLeuPr	2714
QY	607	-----TCTTACGTCGACGATTCACCGGTCGTTATCATCCCTTCC-----	571
Db	2714	AsnThrProThrThrThrIleThrIleThrIleThrHisIleProHisHisHis-ProGlnThrHisL	2734
QY	570	--ATACCAACAAATGAAACAACACAGTTCCAGATGGGGGATATCCGATGCTGTGGACA	513
Db	2734	euleuThrAsnLeuAlaLeuThrThrThrThrIleThrHisProHisHisGly-----T	2751
QY	512	CTGATCATCTCGATTTCCACCTCTGTAAGTTACCATTTGAC-----C	471
Db	2751	HisHisHisHisAsnGlnProHisIleThrHisIleuAspLeuProThrGlyProPheG	2771
QY	470	AGCATTAACAATGGCAATATGCAATCCACATGAT-----GTCTGTGCGATT	426
Db	2771	HisHisHisHisGlyTrpLeuHisSerThr-GlnProGlyAlaGlyAsnValSerAlaAla	2790
QY	425	GGTCTCTCTGTCTGTCGACCAATGACAAGAGGAGCATGATCTCTGGTGCACATTAATCCCA	366
Db	2791	GlyLeu-----AspProThrGlnHisIleProLeuLeuGlyAlaThrIleuGly-Le	2806
QY	365	GACGGTTGGTATGAGATCCCTTTT--GGTCCGCAATCC	329

[illegible]

D	b	1703	AaPrrro-----AaPrrrGlnAsPrThrThrlleuGlulAlalala---	1717
O	y	1178	ATGAAGAAAGAACAACACGGACCAAAAACAACAACAACTGAATTTCGAAATAT	1119
D	b	1717	-----	1717
O	y	1118	CACCTGGCAAAATCCACCCTTTAACSTACACACACAAATAGTGCATACATAC	1058
D	b	1718	--LeuLeuGlnInThrProtyrAlaGlnPro-----	1727
O	y	1058	ATGGCACTATATTCGACGATCAAGTCACTTCAT-----	1023
D	b	1728	----AlaLeuPheAlaPheGlnValAlaLeuIleArgLeuLeuthAsrGlyTrnHs	1745
O	y	1023	-----	1023
D	b	1746	IleThrProHisTyrrylAlaGlyHisSerLeuGlyGluIleThrAlaIleuLeuAla	1765
O	y	1022	-----CTACACACTCTCCAACACATCTCADA	996
D	b	1766	GlyIleLeuthrIleuthrAspAlaThrThrLeuIleThrGlnArgAlaThrMetGln	1785
O	y	995	ACACATGACCCCTACTTGATGACACACACACTTCGATGGTTGCTAATATTTCA	936
D	b	1786	ThrMetProArgLysThrMetThrThrLeuHisThrThr-----	1798
O	y	935	TGAAGTCTCTCTCCCAAGTGTCTCATACCTCCATCAATCCATTC-----	885
D	b	1799	-----ProHisIleThrHisIleThrAlaHisGluHisAspLeu	1813
O	y	884	-----ATCCCATCCAAATGCATCTCACTCAC	858
D	b	1814	AlaIleAlaAlaIleAsnThrProThrSerLeuValIleSerGlyThrProHisThrVal	1833
O	y	857	TCACATACCGTCAAC-----	843
D	b	1834	GlnHisIleThrThrLeuCySGlnGlnGlnIleLeuThrLyThrLeuProThrLyS	1853
O	y	842	-----TCACCTTSCACAAACAATGACCGACGATGAAATCACTCTGCTACAA	795
D	b	1854	AsnAlaPheHisSerProHisThrAsnProIleLeuAsnGlnLeuHisGlnHisThrGln	1873
O	y	794	CTGCACAAATGCTCCAAACAACAGTTGTATATGACAGCTCTGACCAATCAACAAC	735
D	b	1874	ThrLeuthrTyrrHisProProHisIsthrProLeuIleThrAla-----AsnThr-ProPr	1891
O	y	734	GGATCAG-----GTGACACAGCA	717
D	b	1891	AsrGlnLeuLeuthrProHisIstyrThrThrGlnGlnAlaArgAsnThrValAsrTyrrAl	1911
O	y	716	GACACAAACAAG--ATACACAGACACACAGAAATACTGACATCTGACTGACAC	660
D	b	1911	aThrThrThrGlnThrLeuHisGlnHisGly-----ValThr	1223
O	y	659	TGAAGTACGTGCTTATGACCAACAATCACTCTTCTCAATATTCACCAAT-----	608
D	b	1923	rThrTyrrIleGluLeuGlyProAspAsnThrIleuthrThrIleuthrHisAsnLeuPr	1943
O	y	607	-----ICTTAAGTAGACGATCACGGTGTATATCTCCCTTC-----	571
D	b	1943	AsnThrProThrThrThrIleuthrIleuthrHisIsthrProHisIsthrHisIsthr	1963
O	y	570	-ATACCAACAATGAAGAAACACACAGTTCACGTCGCGGAGATTCGATGCGTGGCAGA	513
D	b	1963	euleuthrAsnLeuAlaLythrThrThrThrThrIsthrHisIsthr-----T	1980
O	y	512	CTGATCACTGTGATTCACCTCTGAAGTGCACCAATTTCAGC-----C	471
D	b	1980	hrrHisIsthrHisIsthrProHisIsthrHisIsthrHisIsthrLeuAspLeuProThrTyrrProPheG	2000
O	y	470	AGCATTAACATGCAATATGCAATTCACCTGAC	439

```
Db      2000    lnhslglnhietyrtprleuglueerhrgin 2010
RESULT 5
AAG65265
ID       AAG65265 standard; Protein; 6239 AA.
XX
AC       AAG65265;
XX
DT       04-DEC-2001 (first entry)
XX
DE       Streptomyces avermiltillis protein SEQ ID NO: 5.
XX
KM       Avermectin aglycone synthase, AAS; avermectin derivative;
KW       drug production; veterinary drug; pesticide.
XX
OS       Streptomyces avermiltillis.
XX
PN       WO200162939-A1.
XX
PD       30-AUG-2001.
XX
PF       23-FEB-2001; 2001WO-JP01381.
XX
PR       24-FEB-2000; 2000JP-0047405.
XX
PA       (KITOW ) KYOMA HAKKO KOGYO KK.
XX
PA       (KITA ) KITASATO INST.
XX
PI       Endo H, Yamaguchi H, Kanda Y, Hashimoto S, Omura S, Ikeda H;
DR       WPI; 2001-582053/65.
DR       N-PSDB; AAH79277.
XX
PT       New modified avermectin aglycone synthase derived from Streptomyces
PT       avermiltillis used in production of 22,23-dihydroavermectin B1a used in
PT       drugs and pesticides -
XX
PS       Claim 4; Page 180-201; 257pp; Japanese.
CC
CC       The present invention relates to the production of modified derivatives
CC       of avermectin aglycone synthase (AAS) derived from Streptomyces
CC       avermiltillis. The activity of an acyl carrier protein (ACP),
CC       beta-retalacyl carrier protein synthase (KS), acyltransferase (AT),
CC       beta-retalacyl carrier protein reductase (KR), dehydrogenase (DH), enoyl
CC       reductase (ER) and/or thioesterase (TE) domain may be reduced or
CC       suppressed. The process can be used in the production of drugs, veterinary
CC       drugs and pesticides. The present sequence is an S. avermiltillis protein.
XX
SQ       Sequence      6239 AA;

Alignment Scores:
Pred. No.:          0.000585           Length:         6239
Score:             138.00              Matches:         84
Percent Similarity: 30.34%              Conservative:   41
Best Local Similarity: 20.39%            Mismatches:     117
Query Match:        5.18%               Indels:         170
DB:                22                   Gaps:           17
US-10-020-441-1 (1-1477) x AAG65265 (1-6239)
QY       1358 GGATGCGAGAGAGGTGCACCACTGTCACCTGGAGCGGTTAACTGCCGAATTCTGAC 1299
Dy       1665 GlyMetAlaHisGlyLeuTyrlHisThrHisProValPheAlaIaIaLeuAsn--Aap 1683
QY       1298 CTCGTTTCAACAATAATXAGAATTCGCAATGGTGCGCATGTCCGACATTCCTATTAATGCGCTC 1239
Dy       1684 IleCystrHisrIstleu---AapProHisIstleuAspHisIstleuProIleuProIleuTrnGin 1702
QY       1238 GACCSTGTGGTGCAAAACAACAACAACAACAACAAGTACAGTACAGTACAGTACAGTACAG 1179
Dy       1703 AspPro-----AapTrnGinAspTrnThrThrIleuGlutIaIaIaIaIa---- 1717
QY       1178 ATGAGAAAGAACAAAACAACACGAGAACAAAAACAACAACAAGTAAATTTGAAAATAT 1119
```

```

Db 1717 ----- 1717
QY 1118 CAACGTGCAATATCCACCCTTACACTACCCACACAACATATGCCATACATGAC 1059
Db 1718 ---LeuLeuGlnGlnThrProTyrAlaGlnPro----- 1727
QY 1058 ATGTGCACATATTCGACGAGTCAAGTCACTCTCCAT----- 1023
Db 1728 -----AlaLeuPheAlaPheGlnValAlaLeuHisArgLeuLeuThrAspGlyTyrHis 1745
QY 1023 ----- 1023
Db 1746 IleThrProHisTyrTyrAlaGlyHisSerLeuGlyIleThrAlaAlaHisPheVala 1765
QY 1022 -----CTCACACACTCTCCAAACACATATCTCAA 996
Db 1766 GlyIleLeuThrLeuThrAspAlaThrThrLeuIleThrGlnArgAlaThrLeuMetGln 1785
QY 995 ACACATTCACCCCTACTGATGATGATACACACACACTCTCCATGCTTGCATAGATATTCA 936
Db 1786 ThrMetProProGlyThrMetThrThrLeuHisThrThr----- 1798
QY 935 TGAGTCTTCTCTCCCAAGTGCTCAATCATCCCATCATCAATTC----- 885
Db 1799 -----ProHisHisIleThrHisIleThrAlaHisGlyHisAspLeu 1813
QY 884 -----ATCCCATTCGAATCTACCTCACTCACC 858
Db 1814 AlaIleAlaAlaIleAsnThrProThrSerLeuValIleSerGlyThrProHisThrVal 1833
QY 857 TCACATTCACGTGACC----- 843
Db 1834 GlnHisIleThrThrLeuCysGlnGlnGlnGlyIleLeuThrLeuProThrHis 1853
QY 842 -----TCACTTCACCAACAATGACCGAGATGGAATCATCTCTGTAAGTGA 795
Db 1854 AsnAlaPheHisSerProHisThrAsnProIleLeuAsnGlnLeuHisGlnHisThrGln 1873
QY 794 CTGTGCACAAAGTCATCCAAACACAGTGTATATTCACAGCGTCTGACCAATCAACAACC 735
Db 1874 ThrLeuThrTyrHisProProHisIleThrProLeuIleThrAla-----AsnThr-ProThr 1891
QY 734 GGATCAG-----GTAGACAGACA 717
Db 1891 CAspGlnLeuLeuThrProHisTyrTrpThrGlnGlnAlaArgAsnThrValAspTyrAl 1911
QY 716 GACACACAAACG---ATPACACAGACACAGAAATPACTGACTGACATCTGACGAC 660
Db 1911 aThrThrThrGlnThrLeuHisGlnHisGly-----ValThr 1923
QY 659 TGAGTCAGTGTCTTATTCACCAACACACTTCTTCATATTCACCAT----- 608
Db 1923 ThrTyrTyrIleGlyLeuGlyProAspAsnThrLeuThrThrLeuThrHisAsnLeuDr 1943
QY 607 -----TTTACGTGACGGATCAGCGTCTTATTCACCTCTCC----- 571
Db 1943 CAsnThrProThrThrThrThrLeuThrLeuThrHisIleProHisHisHis-ProGlnThrHis 1963
QY 570 ---ATAACCAACAATGAAAAACACCAAGTTCAGAGTGGCGGATATCCGAGTGGTGGACA 513
Db 1963 euleuThrAsnLeuAlaTyrThrThrThrThrThrProHisHisTyr-----T 1980
QY 512 CTGATCAGCTCTGATTCGACTCTGTAAAGTGCACATTGAGC-----C 471
Db 1980 HisHisHisHisAsnGlnProHisIleThrHisIleLeuAspLeuProThrTyrProPheG 2000
QY 470 AGCATTCATATGCAATATGCAATGCACTGAG 439
Db 2000 InHisGlnHisTyrTrpLeuGlnSerThrGln 2010

```

RESULT 6  
AAB95124

```

ID AAB95124 standard; Protein; 307 AA.
XX
AC AAB95124;
XX
DT 26-JUN-2001 (first entry)
XX
DE Human protein sequence SEQ ID NO:17122.
XX
KW Human; primer; detection; diagnosis; antisense therapy; gene therapy.
XX
OS Homo sapiens.
XX
PN EPI074617-A2.
PD 07-FEB-2001.
XX
PE 28-JUL-2000; 2000EP-0116126.
XX
PR 29-JUL-1999; 99JP-0248036.
PR 27-AUG-1999; 99JP-0300253.
PR 11-JAN-2000; 2000JP-0118776.
PR 02-MAY-2000; 2000JP-0183767.
PR 09-JUN-2000; 2000JP-0241899.
XX
PA (HELI-) HELIX RES INST.
XX
PI Ota T, Ieogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
XX
DR WPI; 2001-318749/34.
XX
PT Primer sets for synthesizing polynucleotides, particularly the 5602
PT full-length cDNAs defined in the specification, and for the detection
PT and/or diagnosis of the abnormality of the proteins encoded by the
PT full-length cDNAs -
XX
PS Claim 8; SEQ ID 17122; 2537pp + CD ROM; English.
XX
CC The present invention describes primer sets for synthesizing 5602
CC full-length cDNAs defined in the specification. Where a primer set
CC comprises: (a) an oligo-dT primer and an oligonucleotide complementary
CC to the complementary strand of a polynucleotide which comprises one of
CC the 5602 nucleotide sequences defined in the specification, where the
CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination
CC of an oligonucleotide comprising a sequence complementary to the
CC complementary strand of a polynucleotide which comprises a 5'-end
CC sequence and an oligonucleotide comprising a sequence complementary to a
CC polynucleotide which comprises a 3'-end sequence, where the
CC oligonucleotide comprises at least 15 nucleotides and the combination of
CC the 5'-end sequence/3'-end sequence is selected from those defined in
CC the specification. The primer sets can be used in antisense therapy and
CC in gene therapy. The primers are useful for synthesizing polynucleotides,
CC particularly full-length cDNAs. The primers are also useful for the
CC detection and/or diagnosis of the abnormality of the proteins encoded by
CC the full-length cDNAs. The primers allow obtaining of the full-length
CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
CC AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to
CC AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632
CC represent oligonucleotides, all of which are used in the exemplification
CC of the present invention.
XX
SQ Sequence 307 AA;

```

Alignment Scores: 0.000296 Length: 307  
 Pred. No.: 136.00 Matches: 85  
 Score: 29.92% Conservative: 32  
 Percent Similarity: 21.74% Mismatches: 112  
 Best Local Similarity: 5.06% Indels: 162  
 Query Match: 22 Gaps: 18  
 Db:

US-10-020-441-1 (1-1477) x AAB95124 (1-307)

```

QY 129 TGTGTTGCTACACATGCGCAGTATTGGGTAAAGATGACGTAAACACATT----- 192
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 12 CysValCysValAlProValCysVal-----GlyAlaHisLeuCysVal 25
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 193 -----GTGCTCAATGTAATGATGATTTCACAGAC-----GTGTTGCGTGGG 240
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 26 TrpValHisIleCysValAlaGlyLeuCysGlyCysThrPheValCysAla 45
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 241 TGGCTTTGTTGTTGTTGATGAGT-----GGGTTTGCATGATTTTAATG 285
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 46 CysValCysAlaCysValCysGlyCysLeuCysValTrpGlyThrPheValCysVal-Cy 65
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 286 TCTATATAGTGGTCTTATTTCAGGTTTCATTTTCACATGAGAAATGCAACA 345
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 65 sAlaCysValTrpValHisIleCysValCysAlaSerValCysValProValCysGly 85
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 346 GGCATTCATCCACACCGCTGAGTTAAGTGGCACCGAGATACAGCCCTTGTATG 405
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 85 yAlaHisLeuCysValAlaGlyLeuCysValGlyValHisIleCysValCysAlaCysValCy 105
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 406 TCGCACACAGAGAGACCAATCCACACAGACATCAGTGCATGATTTGCAATTGTA 465
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 105 s-GlyCysThr-----PheValCysValCysAlaCysVal 116
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 466 ATGCTGGCTCAATGTCATTAAGATGAGTGAATCAGATGATCAGTCCACAGCCA 525
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 116 aLysGly----- 118
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 526 TCGGATATCCGCCACCTGGAACTGGTGTTCATTGTTGTTAAGAAAGGATGATAC 585
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 119 -----Cys----- 119
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 586 GACCGTGATCCGTCAGTAAGATGTGGATATTGAAGAAAGTGTGGTAATA 645
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 120 -----LeuCysValCysG 124
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 646 AACGACATGACTAGCTGCTGCTAGTGCATGATGATGATTTCTGTGTTCTGTATCT 705
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 124 yAlaHisLeuCysValCysValCysLeuCysValGlyAlaHisLeuCysValCysLeuC 144
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 706 GTTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 756
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 144 yValCysAlaCysVal-----TrpGlyCysThrPheValCysValCysAla----- 159
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 757 TTGATTAATAACAAGTGTGGT-----ATGACTTTGACAGTTCAGTGCAGAGATTT 813
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 160 -----CysValTrpGlyCysThrPheValCysValCysAlaCysValCysG 175
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 814 CCATCTCCGTCATTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 873
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 175 yCysThrPheVal-Cys----- 180
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 874 TTGGATGGAGTGAATGTGATGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 933
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 180 ----- 180
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 934 CATGAATATCTATGCAAAACATGGAAGTGT-----TGTTGATCATGAGTGGGGTCA 990
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 181 -----ValCysAlaCysValCysGlyCysThrPheVal 191
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 991 TGTGTTGATGATGTTGTGGAGAGTGGTGAAGATGAGAGTGAATGATTCGTGCAATTA 1050
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 192 CysValCysLeuCysValTrpValHisIle-----CysVal 203
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1051 GTGCATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1095
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 204 -CysValCysLeuCysValTrpValHisIleCysValCysAlaCysValTrpValHisI 223
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1096 AAGGTTGATTTGTCGAGTTGATAT-----TTGCAATTCCTGTGTGTGTTT 1146
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 223 eCysVal---CysValProValCysValCysLeuCysValTrpValHisIleCysValCy 242
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1147 GTTGTGTCGTTGTGTTGTGTTGTGTTGTGTTGTGTTGTGTTGTGTTGTGTTGTG 1206
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

```

```

Db 242 sValCysLeuCysValTrp-----ValHisIleCysValCysValCy 256
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1207 TGTGTTGCTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTT 1231
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 256 sAlaCysValCysValAlaGlyPro 264
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

RESULT 7
ABG14399
ID ABG14399 standard; Protein, 227 AA.
AC ABG14399;
DT 18-FEB-2002 (first entry)
DE Novel human diagnostic protein #14390.
XX Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW food supplement; medical imaging; diagnostic; genetic disorder.
XX Homo sapiens.
OS WO200175067-A2.
PN 11-OCT-2001.
XX 30-MAR-2001; 2001WO-US08631.
PF 31-MAR-2000; 2000US-0540217.
PR 23-AUG-2000; 2000US-0649167.
XX (HYSE-) HYSEQ INC.
PA Dmanac RT, Liu C, Tang YT;
XX WPI; 2001-639362/73.
DR N-PSDB; AAS78586.
XX

PT New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensic, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity -
XX

PS Claim 20; SEQ ID NO 44758; 103bp; English.
XX

CC The invention relates to isolated polynucleotide (I) and
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC and gene mapping, and in recombinant production of (II). The
CC polynucleotides are also used in diagnostics as expressed sequence tags
CC for identifying expressed genes. (I) is useful in gene therapy techniques
CC to restore normal activity of (II) or to treat disease states involving
CC (II). (II) is useful for generating antibodies against it, detecting or
CC quantitating a polypeptide in tissue, as molecular weight markers and as
CC a food supplement. (II) and its binding partners are useful in medical
CC imaging of sites expressing (II). (I) and (II) are useful for treating
CC disorders involving aberrant protein expression or biological activity.
CC The polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensic, gene mapping, identification of mutations in
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. ABG00010-ABG30377 represent novel human
CC diagnostic amino acid sequences of the invention.
CC Note: The sequence data for this patent did not appear in the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX

SQ Sequence 227 AA;

```

Alignment Scores: 0.000514 Length: 227  
Pred. No.: 133.00 Matches: 67  
Score: 133.00 Conservative: 29  
Percent Similarity: 35.29%





QY 502 AGAGTATCAGTCTGCCACAGCCATCGATATCCGCCACCTGGAATGTTTCATT 561  
Db 138 HistHrValThrLeuProAlaSerGluThrPheProGlyMetProCysTrpVal 157  
QY 562 GTTGCTTATGGAAGGATGATACGACCGTGATCCGTACAGTAAGATGGTGAATATTG 621  
Db 158 ThrGlyTrpGlyAspValAlaPasn-----AsnGlyGly----- 168  
QY 622 AAGAAAGTAGTTGTTGG 639  
Db 169 -----CysTrp 170  
RESULT 10  
ABG14053  
ID ABG14053 standard; Protein; 375 AA.  
XX  
AC ABG14053;  
XX  
DT 18-FEB-2002 (first entry)  
XX  
DE Novel human diagnostic protein #14044.  
XX  
KW Human; chromosome mapping; gene mapping; gene therapy; forensic;  
KW food supplement; medical imaging; diagnostic; genetic disorder.  
XX  
OS Homo sapiens.  
XX  
PN WO200175067-A2.  
XX  
PD 11-OCT-2001.  
XX  
PF 30-MAR-2001; 2001WO-US08631.  
XX  
PR 31-MAR-2000; 2000US-0540217.  
PR 23-AUG-2000; 2000US-0649167.  
XX  
PA (HYSE-) HYSEQ INC.  
XX  
PI Drmanac RT, Liu C, Tang YT;  
XX  
DR WPI: 2001-639362/73.  
DR N-PSDB; AAS78240.  
XX  
PT New isolated polynucleotide and encoded polypeptides, useful in  
PT diagnostics, forensics, gene mapping, identification of mutations  
PT responsible for genetic disorders or other traits and to assess  
PT biodiversity -  
XX  
PS Claim 20; SEQ ID No 44412; 103bp; English.  
XX  
CC The invention relates to isolated polynucleotide (I) and  
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,  
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome  
CC and gene mapping, and in recombinant production of (II). The  
CC polynucleotides are also used in diagnostics as expressed sequence tags  
CC for identifying expressed genes. (II) is useful in gene therapy techniques  
CC to restore normal activity of (II) or to treat disease states involving  
CC (II). (II) is useful for generating antibodies against it, detecting or  
CC quantitating a polypeptide in tissue, as molecular weight markers and as  
CC a food supplement. (II) and its binding partners are useful in medical  
CC imaging of sites expressing (II). (I) and (II) are useful for treating  
CC disorders involving aberrant protein expression or biological activity.  
CC The polypeptide and polynucleotide sequences have applications in  
CC diagnostics, forensics, gene mapping, identification of mutations  
CC responsible for genetic disorders or other traits to assess biodiversity  
CC and to produce other types of data and products dependent on DNA and  
CC amino acid sequences. ABG00010-ABG30377 represent novel human  
CC diagnostic amino acid sequences of the invention.  
CC Note: The sequence data for this patent did not appear in the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pct\_sequences.  
XX

SQ Sequence 375 AA;  
Alignment Scores:  
Pred. No.: 0.00134 Length: 375  
Score: 129.50 Matches: 78  
Percent Similarity: 33.43% Conservative: 34  
Best Local Similarity: 23.28% Mismatches: 119  
Query Match: 4.86% Indels: 105  
DB: 22 Gaps: 10  
US-10-020-441-1 (1-1477) x ABG14053 (1-375)  
QY 1296 GCTTTCACATATPAGATTCGCCATTGGTCCGATGTCGATTCATTATAGTGGCTGGA 1237  
Db 2 AlaAlaSerHisIleHisThrThrLeuHisThrProAlaGlnHisValProThrAlaArgArg 21  
QY 1236 CCGTGTGGTGGAACAAACAAACAAACAAACAAACAAACAAACAAACAAACAAACAAACAA 1177  
Db 22 Pro-----AsnArgSerThrThrCysSerThrPro 32  
QY 1176 GAGAAAGAACAAACAAACAAACAAACAAACAAACAAACAAACAAACAAACAAACAAACAA 1117  
Db 33 Gln-----HisThrTrpArgCysAlaAsnIleLeuPro 43  
QY 1116 ACTGGCACAATATCCACCCCTTACATGCCACACACAAATATGTCACATACATACAT 1057  
Db 44 ThrThrThrHisThrThrAsnHisThrHisAlaThrThr----- 58  
QY 1056 GTGCACATATATTCAGCATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 997  
Db 59 -----ThiSerIysThrThrProThrIleHisLeuPro-ThiHisThrArg 73  
QY 996 AACACATTGACCCCTTACTCATGTACACAAACAAACAAACAAACAAACAAACAAACAAACAA 937  
Db 73 GAlaAlaAspProProThrThrHisThrHisGlnAsp----- 84  
QY 936 ATGAGTCTTCTCTCTCCAAAGTGTGATCATGATCCATGCATGCATGCATGCATGCATGCATGC 877  
Db 85 -----AsnHisIleHisAlaThrThr----- 90  
QY 876 CAATGATCTGATACATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 817  
Db 91 ----ThiProIleArgArgProHisHisThrSerThrProHisIleHis----- 106  
QY 816 TGGAAATCACTGTGCTGAACTGAACTGAACTGAACTGAACTGAACTGAACTGAACTGAACTGAA 757  
Db 107 ----TyrAsnSerIleThrThrGlnGlnThrHisAsnThr----- 120  
QY 756 GGGCTGTGACCAATACAAACAAACCGGATCGAGTTCAGACACACACACACACAAACAGATACACA 697  
Db 121 -----HisArgHisAsnHisThrHisIleAspAsnGln 131  
QY 696 GACACACAGAAATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 638  
Db 131 ThrHisSerProThrGlnAlaThrSerHisIleHisThrAlaHisIleAsnAsnThrPro 151  
QY 637 AACCAATCACTTTCTTCAATATATTCACCAATGCTTACGATGACGATGACGATGATGATGATGAT 578  
Db 151 ThrGlnHis-----ThrThrIleThrAspProGlyThrAlaLeuAlaLeu 166  
QY 577 CCGTTCATATACCAACAAATGAAACAAACAAACAAACAAACAAACAAACAAACAAACAAACAAAC 518  
Db 166 rProAlaLeuThrAlaIleArgThrGlnSerAlaTrpAsnGlnHisIleAlaAlaIleThr 186  
QY 517 G----- 517  
Db 186 rHisLeuSerValGlnArgSerArgAlaThrGlnHisGlnLeuSerGlnArgGlyHis 206  
QY 516 -----CAGACTGATCATCTGTGATTCACCTGTGATGATGATGATGATGATGATGATGATGAT 479  
Db 206 sGlnHisHisThrAspArgAlaHisIleThrAspAsnAsnGlnHisIleArgValValArg 226  
QY 478 TTGAGCCAGCATTAACAATTGCAATATGCAATTCACATGATGTGTGTGGATGTGCTTCC 419



```

Db 226 gGlyIaGin--TyrThrSerAsnIysGluHisIysGluGlnLeuSerArgAlaGlyIy 245
Qy 418 TCTGTCGCGCCGACATACAGAGGCGATGATCCGTCGCCAC 376
Db 245 eGlySerAspAlaHisArgGlnGlyProThrThrIySAlaHis 259

RESULT 11
AAB23749
ID AAB23749 standard; Protein; 3972 AA.
XX
AC AAB23749;
XX
DT 10-JAN-2001 (first entry)
XX
DE S. avermectilis avermectin aglycon synthase protein SEQ ID NO:3.
XX
KM Streptomyces avermectilis; avermectin aglycon synthase; biosynthesis;
KM multifunctional enzyme; polyketide; avermectin; veterinary drug;
KM agrochemical.
XX
OS Streptomyces avermectilis.
XX
PN MO200050605-A1.
XX
PD 31-AUG-2000.
XX
PF 23-FEB-2000; 2000MO-JP01041.
XX
PR 24-FEB-1999; 99JP-0046961.
XX
PA (KITA ) KITASATO INST.
XX
PI Omura S, Ikeda H;
XX
DR WPI; 2000-565458/52.
XX
DR N-PSDB; AAA92301.
XX
PT Avermectin aglycone synthase DNA and proteins encoded by all or part of
PT ic for the production of avermectin and its derivatives for drug and
PT agrochemical use -
XX
XX
PS Claim 32; Page 203-223; 314pp; Japanese.
XX
CC The present sequence represents an avermectin aglycon synthase protein.
CC Also described are: (1) polypeptides encoded by all or part of the DNA;
CC (2) expression vectors containing the DNA; (3) host cells transformed by
CC the vectors; (4) preparation of the polypeptides by culture of the
CC transformants; (5) preparation of avermectin aglycon or its derivatives
CC by culture of transformed avermectin-producing microorganisms; and (6)
CC oligonucleotides of 5-60 bases in length containing sense or antisense
CC sequences from the avermectin aglycon synthase DNA. The enzymes are
CC useful for the production of modified forms of avermectin and of the
CC intermediates in its biosynthesis, for use as drugs, veterinary drugs
CC and agrochemicals.
XX
SQ Sequence 3972 AA;

Alignment Scores:
Pred. No.: 0.00403 Length: 3972
Score: 128.50 Matches: 106
Percent Similarity: 37.00% Conservative: 62
Best Local Similarity: 23.35% Mismatches: 181
Query Match: 4.82% Indels: 105
DB: 21 Gaps: 23

US-10-020-441-1 (1-1477) x AAB23749 (1-3972)
Qy 1381 GACACCGAGTACGACCTTGAAGGATGAGAGAGAGTCCACCACTGACCTG---- 1325
Db 1439 GUSerArgPheTyrAspAlaValAlaArgGlnAspLeuThrValAlaThrIleu 1458
Qy 1324 -----AGCGGGTAAGTGTCCGAATTT-----CTGACCTGCTTTCACACA 1286

```

```

Db 1459 AlaValProProSerAlaGlyLeuAspThrValAlaProAlaLeuSerAlaTyrHisArg 1478
Qy 1285 TTAGAGATTGCCATTGGTCGATGTCGGCATTCCTATTATGTCGTCGACCCGTGGTGC 1226
Db 1479 HisGlnHisAspGlnAlaArgIleAsnThrTyrThrTyrGlnGlyThr-----TrrLys 1496
Qy 1225 AAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACA 1175
Db 1497 ProLeuThrIleuProThrThrHisIleGlnProHisIleGlnThrIleAlaIleProGlu 1516
Qy 1174 GAAGAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACA 1115
Db 1517 ThrGlnThrHisIle-----ProHisIleThrAsnIle-----LeuThrAsnLeuHis 1532
Qy 1114 TGGCAATATTCACCCCTTTA-----CACTAACCAACAACAATAGTCA 1070
Db 1533 HisIleGlyIleThrProIleProLeuThrLeuAsnHis-ThrHisIleAsnProGlnHis 1552
Qy 1069 CATACAAATCACAATGCACTATATTCAGAGATGCAAGTCACTGCACTGCACTGCTCC 1010
Db 1552 IleuHisIleThrLeuHisIleThrArgGlnGlnIle-----GlnAsnHisThrHis 1569
Qy 1009 AAACAATATCACAACAACAATGACCCCTACTTCATG---TACACAACAACATTCATCG 953
Db 1569 rGlyAlaIleThrGlyLeuLeuSerLeuAlaLeuAspGluThrProHisIleHis 1589
Qy 952 TTTGCAATGATTTTTCAGAGCTTCTGCTCCGCAAGTGTCAATATCCCAATCAAT 893
Db 1589 sProHisIleThrProThrGlyThrLeuLeuAsnLeuThrLeuThrGlnThr-----HisThr 1607
Qy 892 CACATTCAT-----CCATCAATATCTACTCAATGCACTGCACTGCACTGCACTGCACT 844
Db 1607 rGlnThrHisProProThrProLeuThrTyrAlaThrThrAsnAlaThrThrHisIleThr 1627
Qy 843 -----CTCAGCTGCAACAACAACAATGACCGAGATGGAATGCACTGCTGCTGCAAC 794
Db 1627 AsnAspProLeuThrHisIleProThrGlnAlaGlnThrTrr-GlyLeuAlaArgThrThrL 1647
Qy 793 TGTCAAAAGTATCCAAACACAGTTGTTATATC----- 759
Db 1647 euleGluHisProThrHisIleThrAlaGlyIleIleAspLeuProThrThrProThrProH 1667
Qy 758 -----AAGGCTGTGACCAATACACAACCGGATGAGTACAGACAGAC----- 714
Db 1667 IsThrLeuGlnHisLeuThrGlnThrIleThrGlnProHisIleGlnThrGlnLeuAlaI 1687
Qy 713 -ACAACAACAATACACAACAACAACAACAACAATGACTGACTGACTGACTGACTGAG 656
Db 1687 LeArgThrThrGlyThrHisIleThrArgGlyLeuThrProThrThrLeuThrPro----- 1704
Qy 655 TCATGTCGTTATTCACAACAATGCACTGCTTCTTAATATTCACCAATTTTAAGTAGAG 596
Db 1705 -----ThrHisGln-----ProPro----- 1709
Qy 595 GATCAGGTCGTTATTCATCCCTTTCATACACAACAATGAACACACAGTCCAGGTGGG 536
Db 1710 -----ThProThrProHisGlyT 1716
Qy 535 GG-----ATATCCATGAGCTGTGGACAGATGATCACTGTGATTCACCTGTAAAGTTGA 482
Db 1716 hrThrLeuIleThrGlyThrGlyAlaLeuAlaThrHisIleu---Thr-----His 1732
Qy 481 CCAATTGAGCCAGCATTCATGCAATTCGAATATGCAATTCACAGTGTGTCGTGGATGGTGC 422
Db 1732 sHisLeuThrThrHis-----GlnProThrGlnHisIleu----- 1743
Qy 421 TCTCTGTCGCGCCGACATACAGAGGAGATGATCCGTCGCCACCTTAATCCAGAG 362
Db 1744 -----LeuLeuThrSerArgThrGlyProHisIleThrProHisIleGlnHisIleuThrG 1762
Qy 361 GTTGTGTATGATGAGCTGTGTGTGTCGCAATTCCTGAGTGTGAAGAAATGCACTGAATA 302

```

```

Db 1762 nleuGlnGlnGlyIleHisIleuThrIleThrThrCysAspThrSerAsnPro----- 1780
QY 301 AGACCAAGATATAGACATTAAATCTATGCAACCC-----ACTGCATACAC 254
Db 1781 -AspIleuGlnGlnIleuLeuAsnThrIleProProGlnHisProLeuThrThrVal11 1800
QY 253 AAACCAAAACGACGACGACGCAACACACTGTCTGTAATCAATCAATTACATTGAGCAA 194
Db 1800 eHisThrIleGlyIleLeuAspAspAlaThrIleuThrAsnLeuThrProThrGlnLeuAs 1820
QY 193 CAATGTGTTTCAGTCGATCTTACCCCAATCAC 160
Db 1820 nasnValIleuArgAlaIleHisSerAlaHis 1831

RESULT 12
ID AA65264 standard, Protein, 3972 AA.
AC AA65264;
DT 04-DEC-2001 (first entry)
DE Streptomyces avermectin protein SEQ ID NO: 4.
KM Avermectin aglycone synthase; AAs; avermectin derivative;
KN drug production; veterinary drug; pesticide.
XX Streptomyces avermectilis.
XX OS
XX MO200162939-A1.
XX PN
XX PD 30-AUG-2001.
XX PE 23-FEB-2001; 2001MO-JP01381.
XX PR 24-FEB-2000; 2000JP-0047405.
XX PA (KYOM ) KYOMA HAKKO KOGYO KK.
XX PA (KITA ) KITASATO INST.
XX PI Endo H, Yamaguchi H, Kanda Y, Hashimoto S, Omura S, Ikeda H;
XX DR WPI; 2001-582053/65.
XX DR N-PSDB; AAH79277.
XX PT New modified avermectin aglycone synthase derived from Streptomyces
XX PT avermectilis used in production of 22,23-dihydroavermectin B1a used in
XX PT drugs and pesticides -
XX PS Claim 4; Page 167-180; 257pp; Japanese.
XX CC The present invention relates to the production of modified derivatives
XX CC of avermectin aglycone synthase (AAS) derived from Streptomyces
XX CC avermectilis. The activity of an acyl carrier protein (ACP),
XX CC beta-ketoacyl carrier protein synthase (KS), acyltransferase (AT),
XX CC beta-ketoacyl carrier protein reductase (KR), dehydratase (DH), enoyl
XX CC reductase (ER) and/or thioesterase (TE) domain may be reduced or
XX CC suppressed. The process can be used in the production of drugs, veterinary
XX CC drugs and pesticides. The present sequence is an S. avermectilis protein.
XX SQ Sequence 3972 AA;

Alignment Scores:
Pred. No.: 0 00403 Length: 3972
Score: 128.50 Matches: 106
Percent Similarity: 37.00% Conservative: 62
Best Local Similarity: 23.35% Mismatches: 181
Query Match: 4.82% Indels: 105
Db: 22 Gaps: 23

US-10-020-441-1 (1-1477) x AA65264 (1-3972)
QY 1381 GACACCGAAGTGAAGTGAAGGAGAGAGAGTGCACCACTGTCACTGG--- 1325

```

```

Db 1439 GUserArgPheItrpAspAlaValAlaArgGlnSerLeuGlnThrValAlaIleThrLeu 1458
QY 1324 -----AGCGGGTAACGTGCCGAAT-----CTGACCTGCTTCACACA 1286
Db 1459 AlaValProProSerAlaGlyLeuAspThrValAlaProAlaIleSerAlaIlePheSarG 1478
QY 1285 TATAGATTGCGCATTTGCGCATGTGCGCATTCATTAATGCGTGCACCTGTGTGC 1226
Db 1479 HisGlnHisAspGlnAlaArgIleAsnThrTrpThrTrpGlnGlnThr-----TrpLys 1496
QY 1225 AAACAACACACACACACACACAGTACACAAAC-----GTACAGTACAGATGA 1175
Db 1497 ProLeuThrIleProThrThrHisGlnProHisGlnThrTrpLeuIleAlaIleProGlu 1516
QY 1174 GAAAGACAAACAAACACAGGAAACAAACAAACAAAGTCAATTCGAAATTCAC 1115
Db 1517 ThrGlnThrHisHis-----ProHisIleThrAsnIle-----LeuThrAsnLeuHis 1532
QY 1114 TGGCACAATATCCACCCTTTA-----CACTACCCACACAAATAGTCCA 1070
Db 1533 HisHisGlyIleThrProIleProLeuThrIleuAsnHis-ThrHisThrAsnProGlnHis 1552
QY 1069 CATCAATCAACATGTGCACTATATTCGACGATCAAGTCACTCTCCATCTCACCACTGCC 1010
Db 1552 sIeuHisIleThrLeuHisIleHisIleThrArgGlnGlnAla-----GlnAsnHisIleThr 1569
QY 1009 AAACACATATCAAAACACATTCGACCCCTACTTCAG--TACACACACACTTCATCG 953
Db 1569 rGlyAlaIleThrGlyLeuLeuSerLeuLeuAlaLeuAspGlnThrProHisProHisIle 1589
QY 952 TTTCATAGATATTTTCATGAGCTTCTCTCTCAAGTGCTCAATCAATCCATCAAT 893
Db 1589 sProHisIleThrProThrGlnThrLeuLeuAsnLeuThrIleuThrGlnThr-----HisThr 1607
QY 892 CACATTCAT-----CCATCCATCTACCTCACTCACTCACTCACTCACTCACTCACT 844
Db 1607 rGlnThrHisProProThrProLeuThrTrpAlaThrThrAsnAlaThrThrHisIleThr 1627
QY 843 -----CTCACCTCAACCAACCAATGACCGAGATCAATGCACTGCTGACGAAAC 794
Db 1627 oAsnAspProLeuThrHisIleProThrGlnAlaGlnThrTrp-GlyLeuAlaArgThrThr 1647
QY 793 TGTCAAAAGTCATCCAAACACAGCTGTATTATTC----- 759
Db 1647 euleuGlnHisProThrHisThrIleGlyIleIleAspLeuProThrThrProThrProH 1667
QY 758 -----AAGCTGTGACCAATTAACAACACCGGATCAGTGCAGACAGAC----- 714
Db 1667 iethrLeuGlnHisIleuThrGlnThrIleuThrGlnProHisIleGlnThrGlnLeuAlaI 1687
QY 713 --ACCAAAACATATCCACAGACACACAGAAATACGATGATGATGATGACGACTGAG 656
Db 1687 laArgThrThrIleGlyThrHisIleThrArgLeuThrProThrThrLeuThrPro----- 1704
QY 655 TCATGTCGTTTATTCACCAACAACTCACTTCTTCAATATTCACCACTTCTTAACGAGAG 596
Db 1705 -----ThrHisGln-----ProPro----- 1709
QY 595 GATCAGCGTGTATCATCCCTTCCATTAACAACAATGAAACACCAAGTCCAGGTGGCG 536
Db 1710 -----ThrProThrProHisGlyT 1716
QY 535 GG-----ATATCCAGTGTGTCGAGACGTATCATCTGATTCGATCTGTAAGTTGA 482
Db 1716 hrThrLeuIleThrGlyGlyThrGlyAlaLeuAlaThrHisIleu-Thr-----His 1732
QY 481 CCATTGACCGCAGCATTAATGCAATTCGAATTCGAATTCAGTGTCTGTGCGATTGATGATC 422
Db 1732 sHisIleuThrThrHis-----GlnProThrGlnHisIleu----- 1743
QY 421 TCCTCTGTGTCGCGACATACAGAGGCGATATCTCTGTGCGACCTTAATCTCAGAGCG 362

```

Db 1744 ----LeuLeuThrSerArgThrGlyProHisThrProHisAlaGlnHisLeuThrThrG1 1762  
 Qy 361 GTTGGTGAATGATGCTTGTGGTCCATTCCTGATGTAAGAAATGAACCTGAAATA 302  
 Db 1762 nLeuGlnGlnysGlyLeuHisLeuThrLeThrThrCysAspThrSerAsnPro----- 1780  
 Qy 301 AGACCAAGATATAGACATTAAATCTATGCAAAACCC-----ACTGCATACAC 254  
 Db 1781 -AspGlnLeuGlnGlnLeuLeuAsnThrLeProProGlnHisProLeuThrThrVal11 1800  
 Qy 253 AAACCAAAAGCAGCAGCAGCAACCAACACCTGTGTGAAATCATCATGATTACATTAGCAA 194  
 Db 1800 eHisThrAlaGlyLeuLeuAspAlaThrLeuThrAsnLeuThrProThrGlnLeuAs 1820  
 Qy 193 CAATGTGTTTCACTGCATCTCTTACCCGAAATCAG 160  
 Db 1820 nAsnValLeuArgAlaGlyAlaHisSerAlaHis 1831  
 RESULT 13  
 AAG65268  
 ID AAG65268 standard; Protein; 3972 AA.  
 AC AAG65268;  
 DT 04-DEC-2001 (first entry)  
 DE Streptomyces avermitilis protein derivative SEQ ID NO: 8.  
 KM Avermectin aglycone synthase; AAs; avermectin derivative;  
 KN drug production; veterinary drug; pesticide.  
 OS Synthetic.  
 PN WO200162939-A1.  
 PD 30-AUG-2001.  
 PF 23-FEB-2001; 2001WO-JP01381.  
 PR 24-FEB-2000; 2000JP-0047405.  
 PA (KYOW ) KYOWA HAKKO KOGYO KK.  
 PA (KITA ) KITASATO INST.  
 PI Endo H, Yamaguchi H, Kanda Y, Hashimoto S, Omura S, Ikeda H;  
 DR WPI: 2001-582053/65.  
 DR N-PSDB; AAH79279.  
 PT New modified avermectin aglycone synthase derived from Streptomyces  
 PT avermitilis used in production of 22,23-dihydroavermectin B1a used in  
 PT drugs and pesticides -  
 XX Claim 5; Page 235-248; 257pp; Japanese.  
 CC The present invention relates to the production of modified derivatives  
 CC of avermectin aglycone synthase (AAS) derived from Streptomyces  
 CC avermitilis. The activity of an acyl carrier protein (ACP),  
 CC beta-ketocacyl carrier protein synthase (KS), acyltransferase (AT),  
 CC beta-ketocacyl carrier protein reductase (KR), dehydratase (DH), enoyl  
 CC reductase (ER) and/or thioesterase (TE) domain may be reduced or  
 CC suppressed. The process can be used in the production of drugs, veterinary  
 CC drugs and pesticides. The present sequence is derived from an S.  
 CC avermitilis protein.  
 XX Sequence 3972 AA;  
 SQ  
 Alignment Scores:  
 Pred. No.: 0.00403 Length: 3972  
 Score: 128.50 Matches: 106  
 Percent Similarity: 37.00% Conservative: 62  
 Best Local Similarity: 23.35% Mismatches: 181  
 Query Match: 4.82% Indels: 105

DB: 22 Gaps: 23  
 US-10-020-441-1 (1-1477) x AAG65268 (1-3972)  
 Qy 1381 GACACCGAGTACTGACCTTGAGAGGATGGGAGAGAGGTCCACACCTGTCCAGTGC--- 1325  
 Db 1439 GUSerArgPheTyrAspAlaValAlaArgGlnAspLeuGluThrValAlaThrLeu 1458  
 Qy 1324 -----AGCGGGTAACTGTCCGAAAT-----CTGACCTGTTTCACACA 1286  
 Db 1459 AlaValProProSerAlaGlyLeuAspThrValValProAlaLeuSerAlaTrpHisArg 1478  
 Qy 1285 TATAGATTCGCAATGTGTGCGATGTGCGCATTCATTAATGCTGTGACCTGTGTGC 1226  
 Db 1479 HisGlnHisAspGlnAlaArgGlyLeuAsnThrTrpTrpGlnGluThr-----TrpLys 1496  
 Qy 1225 AAACAAACACCAACACACAGTACAGTACAAAC-----GTACAGTACAGATGA 1175  
 Db 1497 ProLeuThrLeuProThrThrHisGlnProHisGlnThrTrpLeuAlaAlaLeProGlu 1516  
 Qy 1174 GAAAGACAAACAAACACAGGAAACAAACAAACAAACAGTGAATTTGAAATATCAAC 1115  
 Db 1517 ThrGlnThrHisHis-----ProHisLeThrAsnHis-----LeuThrAsnLeuHis 1532  
 Qy 1114 TGGCAAAATATCCACCTTTA-----CACTACCCACACAAATAGTCCA 1070  
 Db 1533 HisHisGlyLeuThrProLeuProLeuThrLeuAsnHis-ThrHisThrAsnProGlnHis 1552  
 Qy 1069 CATACAAATCAGATGTGCACTATATTGACAGATCAAGTCACTCTCCATCTCAGTCTCC 1010  
 Db 1552 sLeuHisHisThrLeuHisHisHisThrArgGlnGlnAla-----GlnAsnHisThrThr 1569  
 Qy 1009 AAACACATATCCAAACACATTTGACCCCTACTCTCATG---TACACACACATTCCTACG 953  
 Db 1569 rGlyAlaLeuThrGlyLeuSerLeuLeuAlaLeuAspGluThrProHisProHisHis 1589  
 Qy 952 TTTCGATGATATTTCAATGAGCTTCTCTCCCAAGTGTCTCAATATCCATCCATACAT 893  
 Db 1589 sProHisThrProThrGlnGlyThrLeuLeuAsnLeuThrLeuThrGlnThr-----HisThr 1607  
 Qy 892 CACATTCAT-----CCATCTCAATTTAAGTCACTCACTCACTCACTCACTCAGTCACTC-- 844  
 Db 1607 rGlnThrHisProProThrProLeuThrProLeuTrpGlyAlaThrThrHisAlaThrThrHisHisBr 1627  
 Qy 843 -----TCCACCTCCACACACAAATGACCGAGATGGAATATCATCTGTGTAAG 794  
 Db 1627 oAsnAspProLeuThrHisProThrGlnAlaGlnThrTrp-GlyLeuAlaArgThrThrL 1647  
 Qy 793 TGTCAAAAGTATCATCAACACAGTTGTATTATC----- 759  
 Db 1647 eLeuGlnHisProThrHisHisThrAlaGlyLeuLeuAspLeuProThrThrProThrProHis 1667  
 Qy 758 -----AAGGCTCTGACCAATACACACACCGGATCAGGTAGACAGACAGAC----- 714  
 Db 1667 sThrLeuGlnHisProThrHisHisThrLeuThrGlnThrProHisHisGlnThrGlnLeuAla 1687  
 Qy 713 --ACCAAAAGATPACAGACACACAAATATGATGATGATGATGATGATGATGATGATG 656  
 Db 1687 leArgThrThrGlyThrHisHisThrArgGlyLeuThrProThrThrLeuThrPro----- 1704  
 Qy 655 TCATGTCTGTTATTCACCAACAACTCACTTCTCAATATTCACCACTTTTACGTCAGG 596  
 Db 1705 -----ThrHisGln-----ProPro----- 1709  
 Qy 595 GATCAGGCTGTATCATCTCTCCATTAACCAACATGAAACACAGTTCGAGGTGCG 536  
 Db 1710 -----ThrProThrProHisGlyL 1716  
 Qy 535 GG-----ATATCCGATGCTGTGCGACATGATCATCTGATTCCTGTAAGTTGA 482  
 Db 1716 hrThrLeuThrGlyGlyThrGlyAlaLeuAlaThrHisHisLeu-----His 1732  
 Qy 481 CCATTGAGCCAGATTAACAATTCGAATATGCAATTCACAGTGTGTGTGCGATTGTC 422

```

Db      1732  SHISLeuThrHis-----GlnProThrGlnHisLeu----- 1743
QY      421  TCCTGTGCGTCCGACATACAGAGGCGATGATCCGTGTCACCTTAACGCCAGACG 362
Db      1744  ----LeuLeuThrSerArgThrGlyProHisThrProHisAlaGlnHisLeuThrGln 1762
QY      361  GTTGGTGATGATGAGCTTTGGTGGCCATTCCTCAGTGTGAGAAATGAAACCTGAATA 302
Db      1762  nLeuGlnGlnysGlyLeuHisLeuThrIleThrThrCysAspThrSerAsnPro----- 1780
QY      301  AGACCAAGATATACATTAATAATCTATGCAAAACC-----ACTGCATACAC 254
Db      1781  -AspGlnLeuGlnGlnLeuLeuAsnThrIleProProGlnHisProLeuThrThrVal11 1800
QY      253  AAACCAAAAGCAGCAGCAGCAGCAACACGTCGTGCAATCAATGATTAATGAGCA 194
Db      1800  eHisThrAlaGlyIleLeuAspAspAlaThrLeuThrAsnLeuThrProThrGlnLeuAs 1820
QY      193  CAATGTGTTTCAGTCGATCTTACCCGAAATCAC 160
Db      1820  nAsnValLeuArgAlaAlaHisSerIleHis 1831
RESULT 14
AAW68540
ID      AAW68540 standard; Protein; 190 AA.
XX      AC      AAW68540;
XX      DT      12-FEB-1999 (first entry)
XX      DE      Amino acid sequence of the trypsin-L gene clone NWB352/3.
XX      KW      Tumour rejection antigen precursor; TRAP; TRH; leukaemia; screening;
XX      KM      lymphoma; cancer; HLA; human lymphocyte antigen; vaccine;
XX      KN      trypsin-L.
XX      OS      Homo sapiens.
XX      PN      MO9849299-A1.
XX      PD      05-NOV-1998.
XX      PF      22-APR-1998; 98WO-US07784.
XX      PR      25-APR-1997; 97US-0845998.
XX      PA      (LUDW-) LUDWIG INST CANCER RES.
XX      PI      Boon-Falleur T, Coulie PG, De Smet C, Lucas S, Van Baren N;
XX      DR      MPI, 1999-009425/01.
XX      DR      N-PSDB; AAV33911.
XX      PT      New diagnosis of leukaemia - by detecting genes for tumour antigen
XX      PS      rejection precursors or corresponding proteins
XX      PS      Claim 11; Page 62; 88pp; English.
XX      CC      The present sequence represents the amino acid sequence of the
XX      CC      trypsin-L gene clone NWB352/3. The protein is a tumour rejection
XX      CC      antigen precursor (TRAP). The specification describes the treatment of
XX      CC      disorders which characterised by expression of a leukaemia-associated
XX      CC      nucleic acid such as TRH. The products are used for in vivo or in vitro
XX      CC      screening for leukaemia, lymphoma or other cancers by usual
XX      CC      hybridisation/amplification or immunoassay methods. TRAPs, when
XX      CC      processed to antigens or complexed with HLA (human lymphocyte antigen)
XX      CC      molecules, or nucleic acid encoding them, are useful in vaccines for
XX      CC      treating leukaemia.
XX      SQ      Sequence 190 AA;

```

Alignment Scores:

```

Pred. No.: 0.00202 Length: 190
Score: 126.50 Matches: 41
Percent Similarity: 34.41% Conservative: 23
Best Local Similarity: 22.04% Mismatches: 45
Query Match: 4.71% Indels: 77
Db: 20 Gaps: 5
US-10-020-441-1 (1-1477) x AAW68540 (1-190)
QY      91  TGATACAGSTTCACSTAGTCTCAACGAGACAGTACTCAGAGTGTGATTTGTT----- 144
Db      59  CysGlyGlySerPheIleHisProGlnIleThrValIleThrAlaAlaHisCysValGlnPro 78
QY      145  ---TCCTCAGCATTCGCCAGTATTCGGGTAAAGATTCAGTGAACACATTTGTTCTGAA 201
Db      79  AspIleLysAspLeuAlaAlaLeuArgValGlnLeuArgGlnGlnHisLeuThrTyrosGln 98
QY      202  TGTAAATCGATTGATTTCACAGACAGTGTTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTC 261
Db      99  AspGln----- 100
QY      262  AGTGGGTTTGCATAGATTTTAAATGCTATATATCTTGTATTTTCAAGTTTCATTTCTC 321
Db      101  -----LeuLeuProValSerArgIleIleValHisProGlnPheTyr 114
QY      332  ACACCTGAGGAATGCGACCAACACAGCATTCATCACAAACGTCGTGAATTAAAGTGGCA 381
Db      115  IleIleGln----- 117
QY      382  CCAGATACATGCCCTCTTTGATGTGCGCAGACAGAGAGACCAATGCCACAGACACTC 441
Db      117  ----- 117
QY      442  AGTGATTCGATATTCATTTGATTTGATTCGCTCAATGTCTCAATGTCTCAATGTCTCAATG 501
Db      118  ThrGlyAlaAspIleAlaLeuLeuGlnGlnIleProValAsnIleSerSerHisIle 137
QY      502  AGAGTATCAGTCTGCCACAGCCATCGATATCCGCCACCTGGAACCTGAGTGTTCATT 561
Db      138  HisThrValThrLeuProArgAlaSerGlnThrPheProGlnIlePheProCysThrVal 157
QY      562  GTTGGTTATGAGAAGGATGATACGACCGGTGATCCGTCACGTAAGAATGTGGAATATTG 621
Db      158  ThrGlyTyrGlyAspValAspAsn-----AsnGlyGly----- 168
QY      622  AAGAAAGTGAATTGTTGG 639
Db      169  -----CysTrp 170
RESULT 15
ABW69806
ID      ABW69806 standard; Protein; 1795 AA.
XX      AC      ABW69806;
XX      DT      26-MAR-2002 (first entry)
XX      DE      Drosophila melanogaster polypeptide SEQ ID NO 36210.
XX      KW      Drosophila; developmental biology; cell signalling; insecticide;
XX      KM      pharmaceutical.
XX      OS      Drosophila melanogaster.
XX      PN      WO200171042-A2.
XX      PD      27-SEP-2001.
XX      PF      23-MAR-2001; 2001WO-US09231.
XX      PR      23-MAR-2000; 2000US-191637P.
XX      PR      11-JUL-2000; 2000US-0614150.
XX

```







PR	06-SEP-2000;	2000US-230865P.
PR	06-SEP-2000;	2000US-230988B.
PR	06-SEP-2000;	2000US-230989P.
PR	06-SEP-2000;	2000US-230990B.
PR	07-SEP-2000;	2000US-230986P.
PR	07-SEP-2000;	2000US-230897P.
PR	07-SEP-2000;	2000US-230951P.
PR	07-SEP-2000;	2000US-231163B.
XX	07-SEP-2000;	2000US-231832P.
PA	(INCY-)	INCYTE GENOMICS INC.
PI	Stuart J., Lincoln SE,	Altus CM, Dufour GE, Chalup MS, Hillman JL;
PI	Monyama AL, Yu JT, Wright RJ,	Gietzen D, Liu TF, Yap PE, Dahl CR;
PI	Montalima MG, Bradley DL,	Rohatgi SD, Harris B, Roseberry AM;
PI	Gerstein EH, Peralta CH,	David NM, Panzer SR, Flores V, Dafio A;
PI	Marwaha R., Chen AJ,	Chang SC, Au AP, Inman RR;
DR	WPI, 2002-315658/35.	
XX	N-PSDB; ABL99895.	
PT	Polynucleotide sequences encoding human secretory proteins useful for	
PT	gene therapy of e.g. genetic deficiency disorders, cancers, and	
PT	diseases caused by intracellular parasites -	
XX	Claim 29; Page 463-464; 585pp; English.	
CC	The invention comprises the amino acid and coding sequences of human	
CC	secretory (SPTM) proteins. The SPTM DNA and amino acid sequences are	
CC	useful for treating a disease or condition associated with the expression	
CC	of functional SPTM. The SPTM DNA sequences are useful for somatic or	
CC	germline gene therapy to correct a genetic deficiency (e.g. severe	
CC	combined immunodeficiency). The SPTM DNA sequences are also useful in	
CC	providing protection against intracellular parasites (e.g. fungal	
CC	parasites and protozoan parasites). The SPTM DNA and protein sequences	
CC	are also useful for diagnosing cell proliferative disorders, cancer,	
CC	immune disorders (e.g. AIDS), neurological disorders (e.g. Parkinson's	
CC	disease), motor neuron disorders, demyelinating diseases (e.g. multiple	
CC	sclerosis), meningitis, abscesses, prion diseases, cerebral palsy,	
CC	neuroskelelral disorders, peripheral nervous system disorders,	
CC	dermatomycosits and polymyositis, myopathy, myasthenia gravis, and mental	
CC	disorders (e.g. Tourette's syndrome). Amino acid sequences ABB97749 -	
CC	ABB97933 represent human secretory proteins of the invention.	
SQ	Sequence    247 AA;	
Alignment Scores:		
Pred. No.:	0.00433	Length:         247
Score:	123.50	Matches:        78
Percent Similarity:	31.65%	Conservative:    41
Best Local Similarity:	20.74%	Mismatches:     91
Query Match:	4.60%	Indels:          167
DB:	23	Gaps:            20
US-10-020-441-1 (1-1477) x ABB97898 (1-247)		
OY    231 TGGGTGCGTGGCTTTGTTGTGTGCATGCGGTTGCATAGATTAAATGCTAT	250	
Db      11 CysValIleValIlyrIeuCysValCysMeclAlaIleuCySlenuSer-----	26	
OY    291 ATACTGGTCTTAATTCACGGTTTCATTCTCACATGAGGAATGGCGACAACAGGCAT	350	
Db      26 -----	26	
OY    351 CCATCAACCAACGCTCTTGGAATTAAAGTGGACCAACAGATAATGCCCTTGTATGTGGC	410	
Db      27 SerCysThrCysValTyValIhIsGlYglYValCysValAlaCys-----	41	
OY    411 AGACACAGAGAACAATCATGCACACATCAGCAGATGGATGATTCGCAT-----TGT	464	
Db      42 -----AlaTrpArgArtyValCysSerPheThrCys	51	
OY    465 NATGTGGCTCAATATGTCACTTAGAGAGTGGAATAGAGTGAATGATGTCACAGCC	524	

[illegible]





Pred. No.: 0.00482 Length: 245  
Score: 123.00 Matches: 38  
Percent Similarity: 35.50% Conservative: 22  
Best Local Similarity: 22.49% Mismatches: 45  
Query Match: 4.58% Indels: 64  
DB: 23 Gaps: 3

US-10-020-441-1 (1-1477) x AAU12023 (1-245)

```
QY 91 TGTACAGSTTCACGTAGTCCAAAGAGAGAGACTCAGACGTGTCATTGTGTT----- 144
DB 29 CysGlyGlySerLeuIleHisProGlnTrpValLeuThrAlaHisCysValGlyPro 48
QY 145 ---TGCTCACCATTCGTCAGTTCGGGTAGAGATTCGATGAAACATTTGTTGCTCAA 201
DB 49 AspValIysAspLeuAlaIleuThrValGlnLeuArgGlnHisLeuValTyrGln 68
QY 202 TGTAAATCGATTGATTCACAGACAGTGTGGTGGTGGCTGGCTTGTGTGTATGC 261
DB 69 AspGln----- 70
QY 262 AGTGGGTTTCATAGATTTTAAATGCTATATCTGCTCTTATTTTCAGTTTCATTCTC 321
DB 71 -----LeuLeuProValSerArgIleIleValHisProGlnPheTyr 84
QY 322 AACTGAGGAATGGCGACCAACAGGCATTCATCAACCGCTCTGAGTTAAGTGGCA 381
DB 85 ThrAlaGln----- 87
QY 382 CCAGGATACATGCCCTCTTGTATGTGCGCAGACAGAGAACCAATCGCACAGACATC 441
DB 87 ----- 87
QY 442 AGTGATTCGATTTGCAATTTGTAATGCTGCTCAATGTCGATTCAGATGGAATC 501
DB 88 IleGlyAlaAspIleAlaLeuLeuGlnLeuGluProValAsnValSerSerHisVal 107
QY 502 AGAGATGATGATGTCGACACAGCCATCGATATCCGCGCACCTGGAACCTGGTTCATT 561
DB 108 HisThrValThrLeuProProAlaSerGluThrPheProProGlyMetProCysTrpVal 127
QY 562 GTTGGTTATGAAAGGATGATTAACGAC 588
DB 128 ThrGlyTyrGlyAspValAspAsnAsp 136
RESULT 21
AAU12024
ID AAU12024 standard; Protein; 245 AA.
XX
AC AAU12024;
XX
DT 09-APR-2002 (first entry)
XX
DE Human beta-II tryptase active site mutant S194A #8.
XX
KW Human; proteolytic trypsin; protease; recombinant beta-II trypsin;
enzyme; mutant; mutein.
XX
OS Homo sapiens.
OS Synthetic.
XX
PN WO200198470-A2.
XX
PD 27-DEC-2001.
XX
PF 20-JUN-2001; 2001WO-US19681.
XX
PR 21-JUN-2000; 2000US-0598982.
XX
PA (PROM-) PROMEGA CORP.
XX
PI Maffei M, Niles AL, Haak-Frendscho M,
XX
```

DR WPI; 2002-114578/15.  
DR N-PSDB; AAS20790.

XX DNA construct for producing enzymatically-inactive proteolytic  
PT trypsin, comprises DNA sequence encoding proteolytic trypsin having  
PT an active site mutation -

XX Claim 40; Page 125-126; 16pp; English.

XX The present invention relates to recombinant human proteolytic  
CC trypsinases, active site mutants of these trypsinases and the methods for  
CC producing these. The method involves the production of a DNA expression  
CC construct comprising a promoter operably linked to a secretion signal  
CC sequence which is operably linked to a DNA sequence encoding a  
CC proteolytic trypsinase with an active site mutation (the construct drives  
CC expression of a mature proteolytic trypsinase that lacks enzymatic activity  
CC due to the active site mutation, in hosts transformed to contain the  
CC construct). The method is useful for producing enzymatically-active  
CC beta-II trypsinase. The active site mutants of proteolytic trypsinase provide  
CC a tool to investigate the structural and functional properties of the  
CC trypsinase and its enzymatic activity, and for modelling studies. The  
CC enzymatically-active, recombinant proteolytic trypsinase produced are  
CC useful as an antigen to generate anti-human trypsinase antibodies  
CC and in drug screening for compounds which act as trypsinase inhibitors,  
CC antagonists, agonists, etc. AAU12009-AAU12024 represent recombinant  
CC human beta-II trypsinase active site mutants.

XX Sequence 245 AA;

XX Alignment Scores:

Pred. No.: 0.00482 Length: 245  
Score: 123.00 Matches: 38  
Percent Similarity: 35.50% Conservative: 22  
Best Local Similarity: 22.49% Mismatches: 45  
Query Match: 4.58% Indels: 64  
DB: 23 Gaps: 3

US-10-020-441-1 (1-1477) x AAU12024 (1-245)

```
QY 91 TGTACAGSTTCACGTAGTCCAAAGAGAGAGACTCAGACGTGTCATTGTGTT----- 144
DB 29 CysGlyGlySerLeuIleHisProGlnTrpValLeuThrAlaHisCysValGlyPro 48
QY 145 ---TGCTCACCATTCGTCAGTTCGGGTAGAGATTCGATGAAACATTTGTTGCTCAA 201
DB 49 AspValIysAspLeuAlaIleuThrValGlnLeuArgGlnHisLeuValTyrGln 68
QY 202 TGTAAATCGATTGATTCACAGACAGTGTGGTGGTGGCTGGCTTGTGTGTATGC 261
DB 69 AspGln----- 70
QY 262 AGTGGGTTTCATAGATTTTAAATGCTGCTTATTTTCAGTTTCATTCTC 321
DB 71 -----LeuLeuProValSerArgIleIleValHisProGlnPheTyr 84
QY 322 AACTGAGGAATGGCGACCAACAGGCATTCATCAACCGCTGGAATTAAAGTGGCA 381
DB 85 ThrAlaGln----- 87
QY 382 CCAGGATACATGCCCTCTTGTATGTGCGCAGACAGAGAACCAATCGCACAGACATC 441
DB 87 ----- 87
QY 442 AGTGATTCGATTTGCAATTTGTAATGCTGCTCAATGTCGATTCAGATGGAATC 501
DB 88 IleGlyAlaAspIleAlaLeuLeuGlnLeuGluProValAsnValSerSerHisVal 107
QY 502 AGAGATGATGATGTCGACACAGCCATCGATATCCGCGCACCTGGAACCTGGTTCATT 561
DB 108 HisThrValThrLeuProProAlaSerGluThrPheProProGlyMetProCysTrpVal 127
QY 562 GTTGGTTATGAAAGGATGATTAACGAC 588
DB 128 ThrGlyTyrGlyAspValAspAsnAsp 136
```





Db 89 Thrla1gin----- 91  
 QY 382 CCAGGATACATGCCCTTGTATGTGCGACGACAGAGACCAATCGCACAGACTC 441  
 Db 91 ----- 91  
 QY 442 AGTGATTCGATATTGCAATTGTAATGCTGCTCAATGCTCACTTACAGATGGAAATC 501  
 Db 92 IlegiyalaAepIleAlaLeuLeuGluLeuGluProValAsnValSerHisVal 111  
 QY 502 AGAGGATCAGCTGCTCCACAGCATCGATATCCGCCACTGGAACGTGTTTCATT 561  
 Db 112 HisThrValThrLeuProProAlaSerGluThrPheProGlyMetProCysTrpVal 131  
 QY 562 GTTGGTTATGGAAGGATGATTAACGAC 588  
 Db 132 ThrGlyTrpGlyAspValAspAsnAsp 140  
 RESULT 25  
 AAM64238  
 ID AAM64238 standard; Protein; 273 AA.  
 AC AAM64238;  
 DT 24-NOV-1998 (first entry)  
 XX  
 DE Human mast cell tryptase I.  
 XX  
 KM Mast cell tryptase I; human; MCP-7; mast cell protease 7;  
 KW blood clot; anticoagulant; myocardial infarction; reocclusion;  
 KW thromboembolism; cerebral embolism; thrombosis; therapy.  
 OS  
 XX Homo sapiens.  
 FH Key Location/Qualifiers  
 FT Peptide 1..18  
 FT /label= Sig\_peptide  
 FT Peptide 19..28  
 FT /label= Pro\_peptide  
 FT Protein 29..273  
 FT /label= Mat\_protein  
 XX  
 XX WC9824886-A1.  
 XX  
 XX PD 11-JUN-1998.  
 XX  
 XX PF 25-NOV-1997; 97WO-US21620.  
 XX  
 XX PR 04-DEC-1996; 96US-0032354.  
 XX  
 XX PA (BGHM ) BRIGHAM & WOMENS HOSPITAL.  
 XX  
 XX PT Stevens RL;  
 XX  
 XX DR WPI; 1998-333308/29.  
 XX  
 XX DR N-PSDB; AAV44329.  
 XX  
 PT New compositions containing tryptase-7, e.g. mouse mast cell  
 PT protease-7 - are used to treat clot formation in e.g. myocardial  
 PT infarction, reocclusion following angioplasty or pulmonary  
 PT thromboembolism  
 PT  
 XX  
 XX PS Disclosure; Page 65; 92pp; English.  
 XX  
 CC This is the deduced amino acid sequence of human mast cell tryptase  
 CC I (see also AAV44329). The invention provides: compositions  
 CC comprising an isolated tryptase-7 that may include chimeric proteins  
 CC that contain (a) a human tryptase for all but the active site region  
 CC and (b) the substrate-binding pocket of mouse tryptase-7 or its  
 CC homologues (see AAM64233-39); a method for treating a blood clot by  
 CC administering a nucleic acid molecule that codes for a tryptase-7,  
 CC or an expression product, to decrease fibrinogen activity; a nucleic  
 CC acid encoding a serine protease (SP); and a method of producing a

CC mature SP by expressing the inactive zymogen in a host cell, and  
 CC cleaving the enterokinase susceptibility domain. The tryptase-7  
 CC polypeptides can be used to treat disorders mediated by undesirable  
 CC thrombus clot formation such as myocardial infarction and  
 CC reocclusion following angioplasty of blood clots associated with  
 CC pulmonary thromboembolism, deep vein thrombosis, cerebral embolism,  
 CC renal vein and peripheral arterial thrombosis. They are also  
 CC useful for all surgical procedures that require decreased blood  
 CC clots.  
 XX  
 SQ Sequence 273 AA;  
 Alignment Scores:  
 Pred. No.: 0.00502 Length: 273  
 Score: 123.00 Matches: 38  
 Percent Similarity: 35.50% Conservative: 22  
 Best Local Similarity: 22.49% Mismatches: 45  
 Query Match: 4.58% Indels: 64  
 DB: 19 Gaps: 3  
 US-10-020-441-1 (1-1477) x AAM64238 (1-273)  
 QY 91 TGTACAGSTTCACTAGTGTCTCAACAGACAGTACTCAAGCTGTTCATTGTT----- 144  
 Db 57 CysGlyGlySerLeuIleHisProGluIntrPValLeuThrAlaIleHisCysValGlyPro 76  
 QY 145 ---TGCTACCATTTGCCAGTATTCGGGTGAAGAGATGCACTGMAACATGTTGCTCAA 201  
 Db 77 AspValIlyAspLeuAlaIleuArgValGlnLeuArgGluGlnHisLeuValTrpGln 96  
 QY 202 TGTATTCATGATTATTCAACAGACAGTGTTCGTCGTCGCTGTTGTTGTGTATGTC 261  
 Db 97 AspGln----- 98  
 QY 262 AGTGGTTTGATAGATTTAATGTCTATTAATCTGCTTATTTCAAGTTCAATTTCTC 321  
 Db 99 -----LeuLeuProValSerArgIleIleValHisProGluInPheTr 112  
 QY 322 ACACAGAGGANTGGCGACCAACAAGCATCCATCAACCGTCGAGTTAAGGTGGCA 381  
 Db 113 Thrla1gin----- 115  
 QY 382 CCAGGATACATGCCCTCTTGTATGTGCGACGACAGAGAGACCAATCGCACAGACTC 441  
 Db 115 ----- 115  
 QY 442 AGTGATTCGATATTGCAATTGTAATGCTGCTCAATGCTCACTTACAGATGGAAATC 501  
 Db 116 IlegiyalaAepIleAlaLeuLeuGluLeuGluProValAsnValSerHisVal 135  
 QY 502 AGAGTATCAGTCTGCCACAGCATCGATATCCGCCACTGGAACGTGTTTCATT 561  
 Db 136 HisThrValThrLeuProProAlaSerGluThrPheProGlyMetProCysTrpVal 155  
 QY 562 GTTGGTTATGGAAGGATGATTAACGAC 588  
 Db 156 ThrGlyTrpGlyAspValAspAsnAsp 164  
 RESULT 26  
 AAM63174  
 ID AAM63174 standard; Protein; 273 AA.  
 AC AAM63174;  
 DT 27-OCT-1998 (first entry)  
 XX  
 DE Human mast cell tryptase I polypeptide.  
 XX  
 KM Mast cell protease; MCP; mouse; inhibitor; peptide substrate; asthma;  
 KM tryptase-6 protein; inflammatory disorder; allergic rhinitis; urticaria;  
 KM antioedema; eczematous dermatitis; atopic dermatitis; anaphylaxis;  
 KW hyperproliferative skin disease; peptic ulcer; hyperresponsiveness;  
 KW inflammatory skin condition; human; mast cell tryptase I.



PI Deleageane AM, Lai P, Hafalia A, Patterson C, Walla NK, Kearney L,  
PI Tribouley CM, Khan FA, Baughn MR, Azimtai Y, Elliott VS,  
PI Nguyen DB, Gandhi AR, Yang J, Hernandez R, Policky JL, Lu DAM,  
PI Reddy R, Yue H, Tang YT;  
XX  
XX WPI: 2002-034518/04.  
DR N-PSDB: AAD23854.

Novel human proteases and polynucleotides encoding the proteases,  
PT useful for treating, diagnosing or preventing cell proliferative,  
PT cardiovascular, autoimmune/inflammatory, neurological and developmental  
PT disorders -

Claim 1; Page 136-137; 151pp; English.

The invention relates to human proteases (PRTS-14) and its corresponding  
CC cDNA molecules. Human PRTS and its nucleic acid molecule are useful for  
CC the diagnosis, treatment and prevention of disorders associated with  
CC increased or decreased expression of PRTS. Examples of such disorders  
CC include, cell proliferative disorders (arteriosclerosis, atherosclerosis,  
CC hepatitis, psoriasis and cancers); autoimmune/inflammatory disorders  
CC (AIDS, Addison's disease, allergy, anaemia, asthma, atopic dermatitis,  
CC diabetes mellitus, glomerulonephritis, multiple sclerosis, osteoporosis,  
CC trauma, Grave's disease, rheumatoid arthritis, ulcerative colitis, and  
CC viral, bacterial, fungal, parasitic, protozoal and helminthic  
CC infections); cardiovascular disorders (myocardial infarction, ischaemic  
CC heart disease and hypertension); neurological disorders (epilepsy,  
CC Alzheimer's disease, Pick's disease, Huntington's disease, dementia,  
CC Parkinson's disease, stroke, mental disorders including mood, anxiety  
CC and seasonal affective disorder and prion diseases); gastrointestinal  
CC disorders (Crohn's disease, anorexia, nausea, diarrhoea and jaundice);  
CC epithelial disorders (contact dermatitis, eczema, acne vulgaris,  
CC alopecia, scabies, insect bites and urticaria); reproductive disorder  
CC (infertility, disruption of estrous and menstrual cycle and  
CC gynaecomastia); and developmental disorders (renal tubular acidosis,  
CC Cushing's syndrome, seizure disorders, congenital glaucoma and cataract).  
CC PRTS DNA is also in useful in gene therapy. PRTS and its immunogenic  
CC fragments are useful for screening libraries of compounds in several drug  
CC screening assays. The present sequence is human protease PRTS-13 protein.  
XX  
XX

Sequence 691 AA:

Alignment Scores:  
Pred. No.: 0.00709 Length: 691  
Score: 123.00 Matches: 38  
Percent Similarity: 35.50% Conservative: 22  
Best Local Similarity: 22.49% Mismatches: 45  
Query Match: 4.58% Gaps: 64  
DB: 23 Indels: 3

US-10-020-441-1 (1-1477) x AAE14348 (1-691)

QY 91 TGTACAGTTCTAGTCTCAACGAGACGACTCAGTGGTTCATTGTT----- 144  
DB 67 CysGlySerLeuIleHisProGlnTrpValLeuThrAlaAlaHisCysValGlyPro 86  
QY 145 ---TGCTACCATTCGCAAGTGTGGGTAAGAATCCGCTAAACACATTGTGCTCA 201  
DB 87 AspValIlyAspLeuAlaIleLeuArgValGlnLeuArgGlnHisLeuTyrTyrGln 106  
QY 202 TGTATTCATTGATTTCACAGACAGTGTGGTGGCGTGGCTTGTGTGTATGATGC 261  
DB 107 AsgGln----- 108  
QY 262 AGTGGTTTGCATGATTATTAATGCTATATACCTGGTCTATTTTCAGTTTCATTCTC 321  
DB 109 -----LeuLeuProValSerArgIleIleValHisProGlnPheTyr 122  
QY 322 ACACAGGAATGGGACCAACAGGCAATCCATCACCAACCGTTCGAGTTAAGTGCGCA 381  
DB 123 ThrIleGln----- 125  
QY 382 CCAGATACATGCCCTCTTGTATGTGCGACGACAGAGAGACCAATCGACAGACATC 441

DB 125 ----- 125  
QY 442 AGTGATTCGATATTGCAATTGTAATGCTGGCTCAATGTGTCACCTACAGATGGAATC 501  
DB 126 IleGlyAlaAspIleAlaIleLeuGlnGluGlnProValAspValSerSerHisVal 145  
QY 502 AGAGTATCAGTCTGCCACAGCAGCATCGATATCCGCCACCTGGAATGTTTTCATT 561  
DB 146 HisThrValThrLeuProProAlaSerGlnThrPheProProGlyMetProCysTrpVal 165  
QY 562 GTTGGTTATGAGGAGATGATACGAC 588  
DB 166 ThrGlyTrpGlyAspValAspAsp 174  
RESULT 28  
AA159288  
ID AA159288 standard; Protein; 957 AA.  
XX  
XX AA159288;  
AC  
XX  
XX 25-APR-2000 (first entry)  
DE Human MUC11 polypeptide.  
XX  
XX Mucin; MUC11; human; chromosome 7q22; epithelial inflammation;  
XX Crohn's disease; ulcerative colitis; asthma; chronic bronchitis;  
XX colorectal cancer; cystic fibrosis; inflammatory bowel disease;  
XX breast cancer.  
XX  
XX Homo sapiens.  
OS  
XX  
XX Key Location/Qualifiers  
FT Modified-site 68 /note="potential N-glycosylation site"  
FT  
XX  
XX MO200004142-A1.  
XX  
XX 27-JAN-2000.  
XX  
XX 16-JUL-1999; 99WO-AU00579.  
XX  
XX 16-JUL-1999; 98AU-0004708.  
XX  
XX (CONN-) COUNCIL QUEENSLAND INST MEDICAL RES.  
XX (ORDE-) ORDER OF SISTERS OF MERCY IN QUEENSLAND.  
XX  
XX Williams SJ, Antalis TM, McGuckin MA, Gotley DC;  
XX WPI: 2000-182416/16.  
DR N-PSDB: AA258824.  
XX  
XX Novel MUC nucleic acid corresponding to mucin gene, useful for treating  
XX PT associated disease conditions e.g. colorectal, breast cancer, cystic  
XX PT fibrosis and inflammatory bowel disease -  
XX  
XX Claim 10; Page 83-88; 103pp; English.  
XX  
XX The invention provides mucin genes (MUC11 and MUC12) located on human  
XX chromosome 7q22. The mucin genes or its portion is used in detecting  
XX CC polymorphism, mutation, deletion, truncation and expansion in the gene  
XX or its gene transcript. Pharmaceutical compositions and gene therapy  
XX constructs comprising the mucin genes are used for treating disease  
XX conditions associated with aberrant Mucin expression, altered properties  
XX of mucus or epithelial inflammatory processes involving Mucins like  
XX CC Crohn's disease, ulcerative colitis, asthma, chronic bronchitis and  
XX CC colorectal cancer, cystic fibrosis, inflammatory bowel disease and breast  
XX CC cancer. The mucin genes and the polypeptides are used for determining  
XX these diseases or their predisposition. The MUC11 and MUC12 polypeptides  
XX are used for preparing antagonist and antibodies. The present sequence  
XX represents the human MUC11 polypeptide.  
XX  
XX Sequence 957 AA;

## Alignment Scores:

Pred. No.:	0.0125	Length:	957
Score:	121.00	Matches:	118
Percent Similarity:	30.29%	Conservative:	38
Best Local Similarity:	22.91%	Mismatches:	179
Query Match:	4.54%	Indels:	180
	21	Gaps:	24

US-10-020-441-1 (1-1477) x AAY59288 (1-957)

```

QY 1366 ACCTTGAAGGATGGAGAGAGGTCACCACTGTCA-----CCTGAGCCGGGTAA 1316
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 344 ThrSerSerGlyValSerGlyLysLeuSerThrThr-SerHisSerArgProGlySerThrhi 363
QY 1315 CTGTCCGAATTCGTACCTGCTTTCACACATATAGATTGCCATTGGTCGATGCGCA 1256
    : ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 363 sThrThrAlaPheProAspSerThrThrThrProGlyLeu-----SerArghi 379
QY 1285 TTCATTT-----ATAGTCGCTCGACCCCTGGTGTGC 1226
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 379 sSerThrThrSerHisSerSerProGlySerThrThrAspThrThrLeuProAlaSerth 399
QY 1225 AAACAACAACCAACAACACAGTACAGTACAAACAAGTACAGTACAGTACAGTAAAGACA 1166
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 399 rThrThrSerGlyProSerGlyLysLeuSerThrThrSerHisSerSerProGlySerThrAs 419
QY 1165 AAACAACAACGGAACAAAACAAAACACACAGTGAATTTGGAAATATCACTGGCAGCAAT 1106
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 419 pThrAlaLeuSerProGlySerThr-----ThrAlaLeuSerPheGlyGlnG1 435
QY 1105 ATCCACCCCTTTAC-----ACTACCCACACACACA----- 1078
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 435 uSerThrThrPheHisSerSerProGlySerThrThrThrThrLeuPheProAspSerth 455
QY 1077 -----ATAGTCACATACATACATACATGTCATATATTCAGCATCAAGTCA 1031
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 455 rThrSerSerGlyLeValGlnAlaSerThrArgValHis-----SerSerThrG1 472
QY 1030 CTCTCCATCTACCACTCTC----- 1010
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 472 ySerProArg-ThrThrLeuSerProAlaSerSerThrSerProGlyLysGlnG1us 492
QY 1009 AAACACATCTCTAAACACATGACCCCTACTTCACTGATACACAAACACTTCATCGTTT 950
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 492 erThrAlaPheGlnThrHisProAlaSerThrHisThrThrProSerThr----- 508
QY 949 GCATGATATTTTCATGATGCTTCTCTCCCAAGTGCTCATATCCCATC----- 897
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 509 -----ProSerThrAlaThrAlaProValGlnG1us 519
QY 896 -----ACATCACATTCATCCATCCATCAATCTACTCACTCACTCACTCACT----- 852
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 519 erThrThrThrHisArgSerProSerSerThrPro-----ThrThrHisPheProAlaS 537
QY 851 --CAGGTCACTCACTCACTCACTCACTCACTCACTCACTCACTCACTCACTCACTCACT 812
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 537 erSerThrThrSerGlyHisSerGlyLysLeuSerThrThrLeuPheHisSerProAspAla- 556
QY 811 ATCACTCTGCTACTGAATGCTGCAAAAGTCATCAAAACAGTTGTTATATATCAAGGCTC 752
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 557 -----SerGlyThrThrProSerSerAlaHis----- 565
QY 751 TGACCAATACACACACCGGATCGGTAGACAGACACACACACACACACACACACACACAC 692
    : ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 566 -----SerThrThrSerGlyArgGlyLysLeuSerThrThrLeuPheHisSerProAspAla- 584
QY 691 ACAGATAACTGACTGACATCTGACTGACGACTGAGTCATGTCGTTTATTCACCAACAC 632
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 584 hGluLeuLeu-----Thrl 589
QY 631 TCACTTTCTTCATATATTCACCATTTCTTACGTACGAGATCACGTCGTTATCATCCCTTC 572
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

```

Db 589 euProGlySerThrThrThrThrProGlyLeuSerGlyAlaSerThrThrPheThrSerSerP 609
QY 571 CATACCAACAATGAAACACACAGTTCAGAGTGGCGGATATCCGATGGCTGTGGCAGAC 512
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 609 roArgSerProThrThrThrThrLeuSerProAlaSerMetThrSerLeuGlyValGlyL- 628
QY 511 TGATCACTGTGATTCACCTGTGTAAGTTGACCATTAGCCAGCATTTCAATCAATAT 452
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 629 -----GluSerThrThrSerArgS 635
QY 451 CGAATCCACTGAGTCTGTGTGGATGTTGCTCTCTGTGTGTGGCAGCATATACAGAGGGA 392
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 635 erGlnProGlySerThr-----HisSerThrV 644
QY 391 TGTATCTGTGGCCACTTAATCCAGCGGTGGTGTGATGCGCTTGTGGTGGCAT 332
    : ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 644 alSerProAlaSerThrThrThrThrProGly----- 653
QY 331 TCCTCAGTGTGAGAAATGAAACCTGAAATTAAGACCAAGTATATAGACATTAATATCTAT- 273
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 654 --LeuSerGlyLysLeuSerThrThr-----ValTyrS 663
QY 272 --GCAACCCCACTGATACACAAACACAAACGACGACACACGCAACCACTGTCTGTGAAA 215
    : ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 663 erSerSerProGlySerThrGlnThr-----ThrVal----- 673
QY 214 TCAATCGATTACATTAGCAACATGTGTTTCAGTCGATCTCTTACCCGAAATCACTGGCA 155
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 674 -----PheProArgSerThrThrThrThrThrValArgG 684
QY 154 ATGTGAGCAACA-----CAATGACAGCTGTGATGACTGCTCTCGTTGAGACTA 104
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 684 lYgluGlnProThrThrPheHisSerArgProAlaSerThrThrThrThrLeuPheThrng 704
QY 103 GTGAASCTGTACACTGTTCTCTCTGTCTGTCTGTCTGTCTGTCTGTCTGTCTGTCTGTCT 44
    : ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 704 lAspSerThrThrSerGlyLeuThrGlnG1usSerThrAlaPheProGlySerProAlaS 724
QY 43 GGTGTTGACAGGTTCAACCACTACGATCAACCAAGTTGAC 3
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 724 erThrGlnThrGlyLeuProAlaThrLeuThrThrAlaAsp 737

```

RESULT 29  
 AAM24513  
 ID AAM24513 standard; Protein; 957 AA.  
 AC AAM24513;  
 XX 12-OCT-2001 (first entry)  
 DE C900P predicted amino acid sequence.  
 XX  
 KW Human; immunotherapy; diagnosis; colon cancer; colon tumour;  
 KW Immunogenic; gene therapy; vaccine; colonic cancer.  
 OS Homo sapiens.  
 PN WO200149716-A2.  
 PD 12-JUL-2001.  
 XX  
 PF 29-DEC-2000; 2000MO-US35596.  
 XX  
 PR 30-DEC-1999; 99US-0476296.  
 PR 10-JAN-2000; 2000US-0480321.  
 PR 15-FEB-2000; 2000US-0504629.  
 PR 06-MAR-2000; 2000US-0519444.  
 PR 19-MAY-2000; 2000US-0575251.  
 PR 29-JUN-2000; 2000US-0609448.  
 PR 28-AUG-2000; 2000US-0649811.  
 XX  
 XX (CORI-) CORIXA CORP.  
 XX







QY 570 ATACCAACATGAAACACGAGTTCAGGTGCGGATATCCGATGCTGTGCGACT 511  
 DB 266 UserProthProSerProProthProAlaProGlnHisSerAspLeuCys----- 283  
 QY 510 GATCAGCTGATTCACCTCTGTAAGTTGACCAATTGAGCAGATTACCAATTGCAATATC 451  
 DB 284 -----LeuAlaMetAlaValG1 289  
 QY 450 GAATCCACTGAGTGTCTGTGCGATTGTT----- 423  
 DB 289 nThrProValProThrAlaAlaGlySerGlyAspArgSerLeuGluAlaLeuG1 309  
 QY 422 ----CTCCCTCTGTGCGGACATACAGAGGCGCATGTATCTCT 384  
 DB 309 yAlaLeuMetAlaAlaLeuAspAspTyrArgGlyGlnPhePro 323  
 RESULT 31  
 AAW64237  
 ID AAW64237 standard; Protein; 275 AA.  
 XX  
 AC AAW64237;  
 DT 24-NOV-1998 (first entry)  
 XX  
 DE Human mast cell tryptase alpha.  
 KM Mast cell tryptase alpha; human; MCP-7; mast cell protease 7;  
 KM blood clot; anticoagulant; myocardial infarction; reocclusion;  
 KM thromboembolism; cerebral embolism; thrombosis; therapy.  
 OS Homo sapiens.  
 XX  
 FH Key Location/Qualifiers  
 FT Peptide 1..20  
 FT /label= Sig\_peptide  
 FT Peptide 21..30  
 FT /label= Pro\_peptide  
 FT Protein 31..275  
 FT /label= Mat\_protein  
 PN W09824886-A1.  
 XX  
 PD 11-JUN-1998.  
 XX  
 PF 25-NOV-1997; 97WO-US21620.  
 XX  
 PR 04-DEC-1996; 96US-0032354.  
 XX  
 PA (BGHM ) BRIGHAM & WOMENS HOSPITAL.  
 XX  
 PI Stevens RL;  
 XX  
 DR WPI: 1998-333308/29.  
 DR N-PSDB; AAW44328.  
 XX  
 PT New compositions containing tryptase-7, e.g. mouse mast cell  
 PT protease-7 - are used to treat clot formation in e.g. myocardial  
 PT infarction, reocclusion following angioplasty or pulmonary  
 PT thromboembolism  
 XX  
 PS Disclosure; Page 63-64; 92pp; English.  
 XX  
 CC This is the deduced amino acid sequence of human mast cell tryptase  
 CC alpha (see also AAW44328). The invention provides: compositions  
 CC comprising an isolated tryptase-7 that may include chimeric proteins  
 CC that contain (a) a human tryptase-7 for all but the active site region  
 CC and (b) the substrate-binding pocket of mouse tryptase-7 or its  
 CC homologues (see AAW64233-39); a method for treating a blood clot by  
 CC administering a nucleic acid molecule that codes for a tryptase-7;  
 CC or an expression product, to decrease fibrinogen activity; a nucleic  
 CC acid encoding a serine protease (SP); and a method of producing a  
 CC mature SP by expressing the inactive zymogen in a host cell, and  
 CC cleaving the enterokinase susceptibility domain. The tryptase-7

CC polypeptides can be used to treat disorders mediated by undesirable  
 CC thrombus clot formation such as myocardial infarction and  
 CC reocclusion following angioplasty of blood clots associated with  
 CC pulmonary thromboembolism, deep vein thrombosis, cerebral embolism,  
 CC renal vein and peripheral arterial thrombosis. They are also  
 CC useful for all surgical procedures that require decreased blood  
 CC clots.  
 XX  
 SQ Sequence 275 AA;  
 Alignment Scores:  
 Pred. No.: 0.0122 Length: 275  
 Score: 119.00 Matches: 36  
 Percent Similarity: 35.50% Conservative: 24  
 Best Local Similarity: 21.30% Mismatches: 45  
 Query Match: 4.43% Indels: 64  
 DB: 19 Gaps: 3  
 US-10-020-441-1 (1-1477) x AAW64237 (1-275)  
 QY 91 TGTACAGSTTCACTAGTCTCAACGAGAGACTCAGAGTGTGATTTGTT----- 144  
 DB 59 CysGlyGlySerLeuIleHisProGlnTrpValLeuThrAlaAlaHisCysLeuGlyPro 78  
 QY 145 --TGCTACCACTTGCACAGTATTCGGTAAAGATCGACTGMAACACATTTGCTCA 201  
 DB 79 AspValIysAspLeuAlaThrLeuNyrValGlnLeuArgGlnGlnHisLeuTyrTrgln 98  
 QY 202 TGTATTCATTTGATTTTCACAGACAGTGTGGGTGGTGGCTTTGGTTTGTGTATGC 261  
 DB 99 AspGln----- 100  
 QY 262 AGTGGTTTGCATAGATTTTAATGCTATATCTGATCTTATTTCAAGTTTCAATTCTC 321  
 DB 101 -----LeuLeuProValSerArgIleIleValHisProGlnPheTyr 114  
 QY 322 ACACAGAGAAATGGCGACCAACAGGCAATCCATCCACCGTGTGAGTTAAGTGGCA 381  
 DB 115 IleIleGln----- 117  
 QY 382 CCAGGATACATGCCCTCTTGTATGTGGCAGCAGACAGAGAACCAATGACACAGACTC 441  
 DB 117 ----- 117  
 QY 442 AGTGATTCGATATTCGATTTGATATGCTGCTCAATATGCTCACTTACAGATGGAATC 501  
 DB 118 ThrGlyAlaAspIleAlaLeuLeuGlnGluProValAsnIleSerSerArgVal 137  
 QY 502 AGAGTATCAGTCTGCCACAGCCATCGGATATCCCGCCACCTGGAACTGTGTTTCATT 561  
 DB 138 HisThrValMetLeuProProAlaSerGluThrPheProProGlyMetProCysTrpVal 157  
 QY 562 GTTGGTTATGGAAGGATGATTAACGAC 588  
 DB 158 ThrGlyTrpGlyAspValAspAsnAsp 166  
 RESULT 32  
 AAW63173  
 ID AAW63173 standard; Protein; 275 AA.  
 XX  
 AC AAW63173;  
 DT 27-OCT-1998 (first entry)  
 XX  
 DE Human mast cell tryptase alpha polypeptide.  
 KM Mast cell protease; MCP; mouse; inhibitor; peptide substrate; asthma;  
 KM tryptase-6 protein; inflammatory disorder; allergic rhinitis; urticaria;  
 KM antiodema; eczematous dermatitis; atopic dermatitis; anaphylaxis;  
 KM hyperproliferative skin disease; peptic ulcer; hypertensive venoses;  
 KM inflammatory skin condition; human; mast cell tryptase alpha.  
 OS Homo sapiens.









XX Human lung mast cell beta-tryptase protein.  
 DE  
 XX  
 KW Beta-tryptase; mast cell; antiallergic; antiinflammatory; antiasthmatic;  
 KW antiarthritis; inflammatory mediator; treatment; diagnosis; dermatitis;  
 KW allergic; inflammatory disorder; tryptase release; allergic bronchitis;  
 KW conjunctivitis; rhinitis; bronchial asthma; autoimmune disease; human;  
 KW arthritis; multiple sclerosis; psoriasis; inflammatory bowel disease;  
 KW tryptase inhibitor; bifunctional.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO9940073-A2.  
 XX  
 PD 12-AUG-1999.  
 XX  
 PF 04-FEB-1999; 99WO-EP00727.  
 XX  
 PR 06-NOV-1998; 98DE-1051300.  
 PR 06-FEB-1998; 98DE-1004761.  
 XX  
 PA (BYKG ) BYK GULDEN LOMBERG CHEM FAB.  
 PA (PLAC ) MAX PLANCK GES FOERDERUNG WISSENSCHAFTEN.  
 XX  
 PI Baer T, Bergner A, Beume R, Bode W, Bundschuh D,  
 PI Dominik A, Goebel K, Huber R, Martin T, Moroder L,  
 PI Pereira PJB, Schaschke N, Sommerhoff C, Stadlweiser J;  
 PI Thibaut U, Ulrich W;  
 XX  
 DR WPI; 1999-479417/40.  
 XX  
 PT New bifunctional inhibitors of tryptase, used for treating and  
 PT diagnosing allergy and inflammation  
 PS  
 PS Disclosure: Fig 20; 26SP; German.  
 XX  
 CC This invention describes novel bifunctional human tryptase (HTA)  
 CC inhibitors (I) of formula K1-L-K2 and their N-oxides (if a heteroaryl,  
 CC heterocycloalkyl or heterocycloalkylene group is present)  
 CC and salts where K1, K2 = head groups, each including a group Q which  
 CC can interact with carboxylate, U = linker of formula -A5-B5-A3-B3-A1-B1-  
 CC M-B2-A2-B4-A4-B6-A6-, which can adopt a conformation such that the K  
 CC groups are 20-45 Angstrom apart and permits (I) to penetrate into a  
 CC volume of 52 x 32 x 40 Angstrom, A1, A2 = CO, NH, O, S, SO2, SO2NH,  
 CC NH-SO2, CONH, NHC(O, COO or direct bond, A3, A4 = CO, CS, S, S, NH,  
 CC OCO, COO, CONH, NHC(O, COO or direct bond, -M-Ra, W = CO or bond, Ra =  
 CC cyclohexylene, phenylene, oxazolidine-2,3-, 3,4- or 3,5-diyl,  
 CC pyrrolidine-1,2- or 1,3-diyl, piperazine-1,4-diyl, morpholine-2,4- or  
 CC 3,4-diyl, thiomorpholine-2,4- or 3,4-diyl or piperidine-1,2-, 1,3- or  
 CC 1,4-diyl, A5, A6 = CO, NH, O, S, CONH, NHC(O, COO or direct bond,  
 CC M = -Ar-; -Ar-CR1R2-Ar-; -Ar-G-Ar-, (cyclohexylene)-E-  
 CC (cyclohexylene) - (where each ring is optionally substituted by 1-3 A');  
 CC 2,5-dioxopiperazine-3,6-diyl; pyridinediyl substituted by R5-R7;  
 CC -(CH2)2CONHCH(COR8)(CH2)4-; 5-8C cycloalkylene (optionally substituted  
 CC by 1-3 A'); or a tricyclic group of formula (i)-(iv) (where each benzo  
 CC ring of (i) and (ii) is optionally substituted by 1-3 A'), Ar =  
 CC phenylene (optionally substituted by 1-3 A'), R1, R2 = H, A' (optionally  
 CC partially or completely substituted by F) or OH, or CR1R2 = 5- or  
 CC 6-membered, optionally substituted cyclic hydrocarbyl; E, T = CH2, O or  
 CC bond; G = O, S or SO2; R5, R6 = H or A'; R7 = H, A', Ph or pyridyl; R8 =  
 CC OA', NH2, NHA', N(A')2, piperidino or morpholino; K1 = -B7(CO)m-B9-X1,  
 CC -B7(CO)m-B9-Y1 or -B7(CO)m-B9-B11-X1; K2 = -B8(CO)p-B10-X2,  
 CC -B8(CO)p-B10-Y2 or -B8(CO)p-B10-B12-X2; B1-B6 = bond or 1-4C alkylene;  
 CC B7-B12 = bond or 1-3C alkylene; m, p = 0 or 1, X1, X2 = NH2,  
 CC -C(NH)-NHRb, guanidino, -O-N=C(NH2)2, -S-C(NH)-NH2, -C(=NH)-SA',  
 CC -NH-C(NH)-C(NH)-NH2 or -C(=N-NH2)-NH2; Rb = H, OH or A';  
 CC A' = 1-4C alkyl; Y1, Y2 = 4-11C heteroaryl or 2-7C heterocycloalkyl,  
 CC containing at least one ring N which can function as proton acceptor  
 CC or donor; Z1, Z2 = 5-12C arylene, 5-12C heteroarylene, 3-8C cycloalkylene  
 CC or 3-8C heterocycloalkylene; arylene, heteroarylene, cycloalkylene,  
 CC 1-heterocycloalkylene, heteroaryl or heterocycloalkyl moieties are  
 CC optionally substituted by 1-3 of OH, halo, NO2, CN, NH2, A', OA', COOA',  
 CC COOA', COOH or CONH2. The products of the invention have antiallergic,  
 CC

CC antiinflammatory, antiasthmatic and antiarthritis activity. HTA is an  
 CC inflammatory mediator released from mast cells. (I) are used for the  
 CC treatment and diagnosis of allergic and/or inflammatory disorders  
 CC associated with tryptase release, e.g. allergic bronchitis,  
 CC conjunctivitis or rhinitis, bronchial asthma, autoimmune diseases such  
 CC as arthritis, multiple sclerosis, dermatitis and psoriasis and  
 CC inflammatory bowel disease. This sequence represents a human lung mast  
 CC cell beta-tryptase which is used to describe the method of the invention.  
 XX  
 SO Sequence 244 AA;  
 SO  
 Alignment Scores:  
 Pred. NO.: 0.0181 Length: 244  
 Score: 117.00 Matches: 37  
 Percent Similarity: 34.91% Conservative: 22  
 Best Local Similarity: 21.89% Mismatches: 46  
 Query Match: 4.36% Indels: 64  
 DB: Gaps: 3  
 US-10-020-441-1 (1-1477) x AAU25925 (1-244)  
 QY 91 TGTACAGSTTCAGTACAGTCAACGAGACGAGTCAAGCTGCTGTTGTT----- 144  
 DB 29 CysGlyGlySerLeuIleHisProGlnTrpValIleuThrAlaIleHisCysValGlyPro 48  
 QY 145 ---TGCCTACCATTCGCCAGTATTCGGGTAAAGATGCACTGAAACACATTGTTGCTCAA 201  
 DB 49 AspValIysAspLeuAlaIleuLysValGlnIleuAspGlnIleuIleuTyrcIln 68  
 QY 202 TGTAACTGATTGATTTCACAGACAGCTTTGGCGTGGCTGTTGTTGTATGTC 261  
 DB 69 Aspgln----- 70  
 QY 262 AGTGGTTTGATAGATTGTTTATGCTATTAATCTTGCTTATTTCAGTTTCATTCTC 321  
 DB 71 -----LeuIleProValSerAlaGlieIleValHisProGlnPheTyr 84  
 QY 322 ACACGTAGGAATGGCGCAACCAACAGCATTCATCAACCGTCTGAGTTAAGGTGCA 381  
 DB 85 ThrAlaGln----- 87  
 QY 382 CCAGATACAGCCCTCTTTGATGTCGGCAGACAGAGAACCAATGCGACACACTC 441  
 DB 87 ----- 87  
 QY 442 AGTGATTGATTCATTCATTCATTCATTCATTCATTCATTCATTCATTCATTCATTC 501  
 DB 88 IleGlyAlaAspIleAlaIleuIleuGluIleuGluProValIleValSerSerHisVal 107  
 QY 502 AGAGTATCAGTCTGCCACAGCATTCGATTCGCCGACCTGGAATCGTGTTCATT 561  
 DB 108 HisThrValThrLeuProProAlaSerGluThrPheProGlyMetProCysTrpVal 127  
 QY 562 GTTGGTTATGCAAGGAATGATTAACAC 588  
 DB 128 ThrGlyTrpGlyAspValAspAsnAsp 136  
 RESULT 38  
 AAU12008  
 ID AAU12008 standard, Protein, 245 AA.  
 XX  
 AC AAU12008;  
 XX  
 DT 09-APR-2002 (first entry)  
 XX  
 DE Recombinant human mature beta-II tryptase.  
 XX  
 DE Human; proteolytic tryptase; protease; recombinant beta-II tryptase;  
 KW enzyme; mutant.  
 XX  
 OS Homo sapiens.  
 OS Synthetic.  
 XX



PN MO200198470-A2.  
 XX 27-DEC-2001.  
 PD  
 XX 20-JUN-2001; 2001WO-US19681.  
 PF  
 XX 21-JUN-2000; 2000US-0598982.  
 PR  
 XX (PROM-) PROMEGA CORP.  
 PA  
 XX Maffei M, Niles AL, Haak-Frendscho M;  
 P1 WPI; 2002-114578/15.  
 DR N-PESDB; AAS20766.  
 XX  
 PT DNA construct for producing enzymatically-inactive proteolytic  
 PT trypsinase, comprises DNA sequence encoding proteolytic trypsin having  
 PT an active site mutation -  
 XX  
 PS Claim 61; Page 80; 126pp; English.  
 CC The present invention relates to recombinant human proteolytic  
 CC trypsinases, active site mutants of these trypsinases and the methods for  
 CC producing these. The method involves the production of a DNA expression  
 CC construct comprising a promoter operably linked to a secretion signal  
 CC sequence which is operably linked to a DNA sequence encoding a  
 CC proteolytic trypsinase with an active site mutation (the construct drives  
 CC expression of a mature proteolytic trypsinase that lacks enzymatic activity  
 CC due to the active site mutation). In hosts transformed to contain the  
 CC construct(s), the method is useful for producing enzymatically-active  
 CC beta-II trypsinase. The active site mutants of proteolytic trypsinase provide  
 CC a tool to investigate the structural and functional properties of the  
 CC process and its enzymatic activity, and for modelling studies. The  
 CC enzymatically-active, recombinant proteolytic trypsinase produced are  
 CC useful as an antigen to generate anti-human trypsin antibodies  
 CC and in drug screening for compounds which act as trypsin inhibitors,  
 CC antagonists, agonists, etc. The present sequence represents recombinant  
 CC human mature beta-II trypsinase.  
 CX  
 SQ Sequence 245 AA:  
 Alignment Scores:  

Pred. No.:	0.0181	Length:	245
Score:	117.00	Matches:	37
Percent Similarity:	34.91%	Conservative:	22
Best Local Similarity:	21.89%	Mismatches:	46
Query Match:	4.36%	Indels:	64
DB:	23	Gaps:	3

US-10-020-441-1 (1-1477) x AMU12008 (1-245)  
 QY 91 TGTACAGSTTCACTACTGTCATCAGACGAGCTACTCACAGTGCGTATTGGTT----- 144  
    |||             |||||             ::             |||||             |||  
 Db 29 CysGlyGlySerleuIleHisProGlnTrpValLeuThrAlaAlaHisCysValGlyPro 48  
                   |||             |||             |||             |||  
 QY 145 ---TGCTCAACATTTGCCAGGATTTGGGTAAGAAGATGCATGAACAACATGTTGCTCAA 201  
                   |||             |||             |||             |||  
 Db 49 AspValLysAspLeuAlaAlaLeuArgValGlnLeuArgLugInhiSbleuTyrryGln 68  
                   |||             |||             |||             |||  
 QY 202 TGTAAATCATGATTTCACACAGACAGTGTTCGCTGCGCTGGCTGTTGTGTGTATGC 261  
    AapGln-----  
 Db 69 AapGln----- 70  
 QY 262 AGTGGGTTTGCAATGATTTTAATGTCATNATACTGSGTCTTATTACAGGTTCAATTCTC 321  
                   |||             |||             |||             |||  
 Db 71 -----LeuLeuProValSerArgIleIleValHisProGlnPhenylr 84  
                   |||             |||             |||             |||  
 QY 322 ACACTGAGGAATGCGACCAACAAGGCATCATCCAACCGTCTGAGATTAAAGTGGCA 381  
    |||             |||             |||             |||  
 Db 85 ThrAlaGln----- 87  
 QY 382 CCAAGATATCATGCGCTCTTGTATGTGCGACAGACAGAGAGACCAATTCGACAGACATCG 441

[illegible]



Search completed: April 9, 2003, 13:30:58  
Job time : 150 secs

---

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.4 p5 4578  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - protein search, using frame\_plus\_n2p model

Run on: April 9, 2003, 13:36:34 ; Search time 42.5 Seconds  
(without alignments)  
4249.314 Million cell updates/sec

Title: US-10-020-441-1  
Perfect score: 2686  
Sequence: 1 gtgtcaacctgtgtacacg.....ctgtacgtcccaatttga 1477

Scoring table: BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Xgapop 10.0 , Xgapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 248812 seqs, 61136040 residues  
Total number of hits satisfying chosen parameters: 497624

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Command line parameters:  
-MODEL=frame+ n2p model -DEV=xlp  
-Q=/cgn2\_1/USFO.spool/US10020441/runat\_09042003\_121214\_1452/app.query.fasta\_1.1671  
-DB=published Applications AA -QFMT=faetan -SUFFIX=rapb -MINMATCH=0.1  
-LOOPCL=0 -LOOPEXT=0 -UNITS=bites -START=1 -END=1 -MATRIX=D1osum62  
-TRANS=human40.cdd -LIST=45 -DOCALLIGN=200 -THR SCORE=pct -THR MAX=100  
-THR MIR=0 -ALIGN=40 -MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEADSIZE=500 -MINLEN=0  
-MAILEN=2000000000 -USER=US10020441 @CGN 1.1.18 @runat\_09042003\_121214\_1452  
-NCU=6 -ICPU=3 -NO XLPXY -NO MMAP -LARGEQUERY -NEG SCORES=0 -WAIT -LONGLOG  
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -Fgapop=6  
-Fgapext=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Published Applications AA.\*

- 1: /cgn2\_6/ptodata/2/pubpaa/US06\_NEM\_PUB.pep.\*
- 2: /cgn2\_6/ptodata/2/pubpaa/PCT\_NEM\_PUB.pep.\*
- 3: /cgn2\_6/ptodata/2/pubpaa/US06\_NEM\_PUB.pep.\*
- 4: /cgn2\_6/ptodata/2/pubpaa/US06\_PUBCOMB.pep.\*
- 5: /cgn2\_6/ptodata/2/pubpaa/US07\_NEM\_PUB.pep.\*
- 6: /cgn2\_6/ptodata/2/pubpaa/US07\_PUBCOMB.pep.\*
- 7: /cgn2\_6/ptodata/2/pubpaa/PCTIS\_PUBCOMB.pep.\*
- 8: /cgn2\_6/ptodata/2/pubpaa/US08\_PUBCOMB.pep.\*
- 9: /cgn2\_6/ptodata/2/pubpaa/US09\_NEM\_PUB.pep.\*
- 10: /cgn2\_6/ptodata/2/pubpaa/US09\_PUBCOMB.pep.\*
- 11: /cgn2\_6/ptodata/2/pubpaa/US10\_NEM\_PUB.pep.\*
- 12: /cgn2\_6/ptodata/2/pubpaa/US10\_PUBCOMB.pep.\*
- 13: /cgn2\_6/ptodata/2/pubpaa/US60\_NEM\_PUB.pep.\*
- 14: /cgn2\_6/ptodata/2/pubpaa/US60\_PUBCOMB.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match Length	ID	Description
1	552	20.6	9	US-10-020-441-2
2	121	4.5	9	US-10-025-380-1065
3	121	4.5	10	US-09-922-217-1065
4	121	4.5	10	US-09-833-263-1065

c	5	120	4.5	196	9	US-09-989-920-224	Sequence 224, App
c	6	119	4.4	275	9	US-10-117-323-36	Sequence 36, App1
c	7	118	4.4	270	10	US-09-923-779-152	Sequence 152, App
c	8	116.5	4.3	273	10	US-09-925-297-695	Sequence 695, App
c	9	111	4.1	276	9	US-10-117-323-35	Sequence 35, App1
c	10	108.5	4.0	272	9	US-10-117-323-40	Sequence 40, App1
c	11	107.5	4.0	153	9	US-10-083-357-1014	Sequence 1014, Ap
c	12	104.5	3.9	5179	9	US-10-025-380-1068	Sequence 1068, Ap
c	13	104.5	3.9	5179	10	US-09-922-217-1068	Sequence 1068, Ap
c	14	104.5	3.9	5179	10	US-09-833-263-1068	Sequence 1068, Ap
c	15	103	3.9	1169	10	US-09-801-368-106	Sequence 106, App
c	16	102.5	3.8	222	9	US-10-012-886-479	Sequence 479, App
c	17	102.5	3.8	222	9	US-09-895-793-479	Sequence 479, App
c	18	102.5	3.8	222	9	US-09-895-814-479	Sequence 479, App
c	19	102.5	3.8	222	10	US-09-759-143-479	Sequence 479, App
c	20	102.5	3.8	222	10	US-09-780-669-479	Sequence 479, App
c	21	102.5	3.8	222	10	US-09-822-827-479	Sequence 479, App
c	22	102.5	3.8	288	10	US-09-216-393-341	Sequence 341, App
c	23	102.5	3.8	288	10	US-09-216-393-344	Sequence 344, App
c	24	102.5	3.8	292	10	US-09-864-761-37944	Sequence 37944, A
c	25	101.5	3.8	692	10	US-09-801-368-274	Sequence 274, App
c	26	100.5	3.7	238	9	US-10-067-761-19	Sequence 19, App1
c	27	100.5	3.7	238	10	US-09-804-156-19	Sequence 19, App1
c	28	99	3.7	272	9	US-10-117-323-37	Sequence 37, App1
c	29	99	3.7	290	9	US-10-041-006A-7	Sequence 7, App11
c	30	99	3.7	290	9	US-10-028-072-222	Sequence 222, App
c	31	99	3.7	290	9	US-10-121-049-222	Sequence 222, App
c	32	99	3.7	290	9	US-10-123-904-222	Sequence 222, App
c	33	99	3.7	290	9	US-10-140-470-222	Sequence 222, App
c	34	99	3.7	290	9	US-10-175-746-222	Sequence 222, App
c	35	99	3.7	290	9	US-10-176-918-222	Sequence 222, App
c	36	99	3.7	290	9	US-10-176-921-222	Sequence 222, App
c	37	99	3.7	290	9	US-10-137-865-222	Sequence 222, App
c	38	99	3.7	290	9	US-10-140-474-222	Sequence 222, App
c	39	99	3.7	290	9	US-10-142-431-222	Sequence 222, App
c	40	99	3.7	290	9	US-10-143-114-222	Sequence 222, App
c	41	99	3.7	290	9	US-10-140-002-222	Sequence 222, App
c	42	99	3.7	290	9	US-10-142-419-222	Sequence 222, App
c	43	99	3.7	290	9	US-10-123-282-222	Sequence 222, App
c	44	99	3.7	290	9	US-10-143-423-222	Sequence 222, App
c	45	99	3.7	290	9	US-10-121-050-222	Sequence 222, App

ALIGNMENTS

RESULT 1  
US-10-020-441-2  
; Sequence 2, Application US/10020441  
; Publication No. US20020182224A1  
; GENERAL INFORMATION:  
; APPLICANT: Doenhoff, Michael  
; APPLICANT: Sayers, Don  
; TITLE OF INVENTION: SCHISTOSOMIASIS VACCINE  
; FILE REFERENCE: Doenhoff  
; CURRENT APPLICATION NUMBER: US/10/020, 441  
; CURRENT FILING DATE: 2001-12-18  
; PRIOR FILING DATE: RPTOR FILING DATE: 1999-10-07  
; NUMBER OF SEQ ID NOS: 3  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 2  
; LENGTH: 106  
; TYPE: PRT  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Description of Artificial Sequence: protein  
US-10-020-441-2  
Alignment Scores:  
Pred. No.: 9.71e-47  
Score: 552.00  
Percent Similarity: 100.00%  
Best Local Similarity: 100.00%  
Length: 106  
Matches: 106  
Conservative: 0  
Mismatch: 0

Query Match: 20.55% Indels: 0  
DB: 9 Gaps: 0  
US-10-020-441-1 (1-1477) x US-10-020-441-2 (1-106)  
QY 310 GTTTCATTTCTCAGCTAGAGATGGCGACCAAGAGCATTCATCACCAGCTGTGGA 369  
DB 1 ValSerPheLeuThrLeuAlaGsnGlyAspGlnGlnGlyLeuIleHisGlnProSerGly 20  
QY 370 GTTAGGGGACCGACGATACATGCGCTTGTATGTGGCGACGACAGAGGACCAATC 429  
DB 21 VallysValAlaProGlyIlyrMetProSerCysMetSerAlaArgGlnArgProIle 40  
QY 430 GCACAGACACTCAGTGGATTCCGATTTGCAATTGTAATGCTGCTCAATGGTCACTTA 489  
DB 41 AlaGlnThrLeuSerGlyPheAspIleAlaIleValMetLeuAlaGlnMetValMetLeu 60  
QY 490 CAGAGTGAATCAGATGATGATCTGTCACAGCCATCGGATATCCCGCACCCTGGACT 549  
DB 61 GlnSerGlyIleArgValIleSerLeuProGlnProSerAspIleProProGlyThr 80  
QY 550 GGTGTTTCATGTGTGTTATGAGAGGATGATACGACCGATCCGTCACGTAAGAT 609  
DB 81 GlyValPheIleValGlyIlyrGlyArgAspAspAspAspArgAspProSerArgLysAsn 100  
QY 610 GGTGGAATTTGAGAAA 627  
DB 101 GlyGlyIleLeuLys 106  
RESULT 2  
US-10-025-380-1065  
Sequence 1065, Application US/10025380  
Publication No. US20020182191A1  
GENERAL INFORMATION:  
APPLICANT: Xu, Jiangchun  
APPLICANT: Lodes, Michael J.  
APPLICANT: Secrist, Heather  
APPLICANT: Benson, Darin R.  
APPLICANT: Meagher, Madeleine Joy  
APPLICANT: Stolk, John A.  
APPLICANT: Wang, Tongtong  
APPLICANT: Wang, Yugu  
APPLICANT: Smith, Carole L.  
APPLICANT: King, Gordon E.  
APPLICANT: Wang, Aijun  
APPLICANT: Clapper, Jonathan D.  
APPLICANT: Skeiky, Yasir A. W.  
APPLICANT: Fanger, Gary R.  
APPLICANT: Vedvick Thomas S.  
APPLICANT: Carter, Darlick  
TITLE OF INVENTION: COMPOUNDS FOR IMMUNOTHERAPY AND DIAGNOSIS  
TITLE OF INVENTION: OF COLON CANCER AND METHODS FOR THEIR USE  
FILE REFERENCE: 210121.471C14  
CURRENT APPLICATION NUMBER: US/10/025,380  
CURRENT FILING DATE: 2001-12-19  
NUMBER OF SEQ ID NOS: 1129  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 1065  
LENGTH: 957  
TYPE: PRT  
ORGANISM: Homo sapiens  
US-10-025-380-1065  
Alignment Scores:  
Pred. No.: 0.00225 Length: 957  
Score: 121.00 Matches: 118  
Percent Similarity: 30.29% Conservative: 38  
Best Local Similarity: 22.91% Mismatches: 179  
Query Match: 4.54% Indels: 180  
DB: 9 Gaps: 24  
US-10-020-441-1 (1-1477) x US-10-025-380-1065 (1-957)

QY 1366 ACCTTGAAGGATGGAGAGAGAGTCCACACTGTCA-----CTGGAAGCGGGTAA 1316  
DB 344 ThrSerSerGlyValSerGluGlnSerThrThr-SerHisSerArgProGlySerThrHi 363  
QY 1315 CTGTCCGAAATTTCTGACCTGCTTTCACACATATAGGATTCGCATTGGTCGATGTCGGA 1256  
DB 363 sThrThrAlaPheProAspSerThrThrProGlyLeu-----SerArgHi 379  
QY 1255 TTCATT-----ATACTGCTCGAACCCCTGTGTGTC 1226  
DB 379 sSerThrThrSerHisSerSerProGlySerThrAspThrThrLeuProAlaSerTh 399  
QY 1225 AAACAACAACCAACAACAAGTACAGTACAAACAAGTACAGTACAGTACAGTACAGTAC 1166  
DB 399 rThrThrSerGlyProSerGlnGlnSerThrThrSerHisSerSerProGlySerThrAs 419  
QY 1165 AAACAACAACGAAACAACAACAACAAGTACAGTACAGTACAGTACAGTACAGTACAGTAC 1106  
DB 419 pThrAlaLeuSerProGlySerThr-----ThrAlaLeuSerPheGlyGlnG1 435  
QY 1105 ATCCACCCCTTAC-----ACTACCCACACACA----- 1078  
DB 435 uSerThrThrPheHisSerSerProGlySerThrHisThrThrLeuPheProAspSerTh 455  
QY 1077 -----ATAGTCCACATACATACATACATGCACTATATTCAGACATCAAGTCA 1031  
DB 455 rThrSerSerGlyIleValGlnAlaSerThrArgValHis-----SerSerThrG1 472  
QY 1030 CTCTCCATCTACCACTCTC-----C 1010  
DB 472 ySerProArg-ThrThrLeuSerProAlaSerSerThrSerProGlyLeuGlnGlyLus 492  
QY 1009 AAACAACATCTCAACAACATTCAGCCCTACTTCATGTAACAACAACATTCATCGTTT 950  
DB 492 eThrAlaPheGlnThrHisProAlaSerThrHisThrThrProSerThr----- 508  
QY 949 GCATGATATTCATGAGCTTCTCTCTCCAGTGGTCTCAATGATCCCATC----- 897  
DB 509 -----ProSerThrAlaThrAlaProValGlnGlnus 519  
QY 896 -----ACATCACTTCATCCATCCATTCATCCATCCATCCATCCATCCATCCATCCAT 852  
DB 519 eThrThrThrThrHisArgSerProSerSerThrPro-----ThrThrHisPheProAla 537  
QY 851 --CAGTCACCTCACCCTCAACAA-----CACATGACCGAGATGGA 812  
DB 537 eSerThrThrSerGlyHisSerGlyLysSerThrThrLeuPheHisSerSerProAspAla 556  
QY 811 ATCACTGTGCTAGTGAAGTGTCAAAAGTCAAAAGTCAAAAGTGTATTTATCAAGGCTC 752  
DB 557 -----SerGlyThrThrProSerSerAlaHis----- 565  
QY 751 TCACCAATTCACAACCGGATGAGTACAGACAGACAGACACAACAACAGATACAGACAC 692  
DB 566 -----SerThrThrSerGlyArgGlyLysSerThrThrSerArgLysSerProGlySer 584  
QY 691 ACAGATATACCTGACATGACATGCTGACGAGTACGATGCTGTATTATCCAAACAC 632  
DB 584 hGlnIleThr-----ThrL 589  
QY 631 TCACCTTCTTCATATTCACCATTTCTTACGTCGATCAGTACGTCGTTATCATCCCTTC 572  
DB 589 euProGlySerThrThrThrProGlyLeuSerGluAlaSerThrThrPheThrSerSerP 609  
QY 571 CATTAACAACAATGAAACACAGTTCACAGTGGGGGATATCCATGGCTGTGCAAGC 512  
DB 609 rArgSerProThrThrThrLeuSerProAlaSerMetThrSerLeuGlyValGlyGlu 628  
QY 511 TGATCACTGTATTCACCTGTAAAGTTGACATTTGAGCCAGATTCATCAATCAATAT 452  
DB 629 -----GlnSerThrThrSerArgS 635  
QY 451 CGAATCACTAGTGTCTGTGGAGATTGTTCTCTGTGTCGACGACATTAAGAAGGGCA 392

```
Db      635   erGlnProGlySerThr-----HisSerThrv       644
Qy      391   TGNATCTGGTGCCACCTTAATCCAGACGGGTGGTGAATGATGCCTTTGGTGCGCAT    3322
Db      644   alserProalaserThrThrThrProgly-----                    653
Qy      331   TCCTCAGCTGTAGAATAAACCCTGMAATTAAACCAAGATATAGACTTTAAATCCTAT-    2737
Db      654   --LeuSerGluGluSerThrThr-----ValTyrS     663
Qy      272   --GCMAACSCCATGTGATCACAAAACAAGACGACGACACGCAAAACATGCTGTGAAA    215
Db      663   erSerSerProGlySerThrGluThr-----ThrAl-----        6737
Qy      214   TCMAATGCATTACATTGAGCAACAATGTGTTTCAGTCGATCTTTAACGCCGAATCACTGCA    158
Db      674   -----pheProArgSerThrThrThrSerValArg     688
Qy      154   ATGCGTAGCAAAACA-----CAATGACCAAGCTGTGAGTACTGCTCGTAGACATA    104
Db      684   lylGluGluProThrThrPheHisSerArgProAlaserThrIleThrThrLuePheThrg    704
Qy      103   GTGAASCTGTACACATTGTCTCTCTGCTGCTTAAGAAATGCSATGAACGGAAATCAAGTC    44
Db      704   luAepSerThrThrSerGlyLeuThrGluGluSerThrAlaPheProGlySerProAlas    724
Qy      43   GGTGTGCACAGGTTACCACTACAGATCAACACAGGTTGAC    3
Db      724   ertHrGlntHrGlyeudProAlaThrIleutHrThrAlaSer    737
```

### RESULT 3

US-09-922-217-1065  
; Sequence 1065, Application US/09922217

; Patent No. US2002007641A1  
; GENERAL INFORMATION:

```
; APPLICANT: Xu, Jiangchun
; APPLICANT: Lodes, Michael J.
```

```
; APPLICANT: Secrist, Heather
; APPLICANT: Benson, Darin R.
```

APPLICANT: Meagher, Madeleine Joy  
APPLICANT: Stolk, John A.

APPLICANT: Wang, Tongtong  
; APPLICANT: Jiang, Yuguang

APPLICANT: Smith, Carole Lynn  
APPLICANT: King, Gordon E.

APPLICANT: Wang, Aijun

1. TITLE OF INVENTION: COMPOUNDS FOR IMMUNOTHERAPY AND DIAGNOSIS

FILE OF INVENTION: OF COLON CANCER AND  
FILE REFERENCE: 210121.471C13  
;  
; CUMMINS ADDITION NUMBER: 110/00/000 017

```

; CURRENT APPLICATION NUMBER: 05/09/922,211
; CURRENT FILING DATE: 2001-08-03
; NUMBER OF SEQ. ID NOS: 1104

```

```

; NUMBER OF SEQ ID NOS: 1124
; SOFTWARE: FastSEQ for Windows Version 4.0.0

```

```

; SEQ ID NO 1065
;
; LENGTH: 957

```

```

; TYPE: PRT
; ORGANISM: Homo sapiens

```

US-09-922-217-1065

Alignment Scores:	
Pred. No.:	0.00225
Length	

Score:	121.00	Match
Percent Similarity:	30.29%	Conse

Best Local Similarity:	22.91%	Mismatch
Query Match:	4.54%	Index

DB:	10	Gaps:
-----	----	-------

US-10-020-441-1 (1-1477) x US-09-922-217-1065 (1-957)

QY 1366 ACCTTGAAGGATGGAGAGAGGTCACCACTGTCA-----CCTGAGCCGGTAA 1316

```

Db      344  Thrsrsergylalsrugiuserthrthr-SerHisserArgProglyserThrHi 363
Qy      1315  CTGCGCAAAATTCTGACCTGCTTTCACACATATAGATTGCCATTGGTCGACATGTCGGCA 1255
Db      363  sThrThrAlaPheProaspserThrThrThrProglyLeu-----SerArgHi 379
Qy      1255  TTTCATT-----ATAGTCGCTGCAGCCCTGGTGGTC 1226
Db      379  sSerThrThrSerHisSerSerProglySerThrThrThrLeuAlaProAlaSerTh 399
Qy      1225  AACAACAACAACCAACAACAGTACAGTACAACAGTACAGTACAGTACAGTACAGTACAGT 1166
Db      399  rThrThrSerGlyProserGlnIuserThrThrHisHisSerSerProglySerThrAs 419
Qy      1165  AAMACAACAGCGAACAAAACAAAACACACAAAGTATTTCGAAATATCACTGGCACAAT 1106
Db      419  pThrAlaLeuSerProglySerThr-----ThrAlaLeuSerPheGlnGlyGlnI 435
Qy      1105  ATCACCCTTTAC-----ACTACCCACACACAA----- 1078
Db      435  uSerThrThrPheHisSerSerProglySerThrThrThrThrLeuPheProaspSerTh 455
Qy      1077  -----ATTGCCACATACAAATCACATGAGCTATATTCAGCATCAAGTCA 1031
Db      455  rThrSerSerGlyTlleAlaAlaIuAlaSerThrArgValHis-----SerSerThrGln 472
Qy      1030  CTCCTCATCTCACACACTCTC-----SerSerThrGln 492
Db      472  ySerProArg-ThrThrLeuSerProAlaSerSerThrProglyLeuGlnGlyGln 492

```

          

Db 492 erthAlapheGInThrHisProAlaserThrhIsThrThrProSerThr----- 508

QY 949 GCATAGATATTGATGAGTCTTCCTCCAGTGGTCTCATCATCCAC----- 897

Db 509 -----ProserThralaProValgiuglus 519

Qy 896 -----ACATCACATTCCATCCCATTCCAATCTACTCCTCACTGACCTGCACAT----- 852

Db 519 erThrThrTyrrHisArgSerProSerSerThnPro-----ThrThrHisPheProAla 537

Qy 851 --CACGTACCTCAGCTCACCAG-----CACAAATGACCGAGATGGAA 812

Db 537 ~~er~~erThrThrSerGlyHisSerGlySerThrIlePheHisSerSerProAspAla- 556

811 ATCACTCTGCTACTGACTGTCAACAAGTCATCCACAACACAGTTGTTATTATCAAGCTC 752

```

Db      557 -----SerGlyThrProSerAlaHis----- 565
          |||  |||  |||  |||  |||  |||

```

751 TGACCAATACAAACCGGATCAGGTAGACAGACAGACACACAAACAGATACACAGACAC 692

```
Db 566 -----SerThrThrSerGlyArgGlyGluSerThrThrThrSerArgIleSerProGlySerT 584
```

691 ACAGATACTGACTGACATCTGACTGACGACTGAGTCATGCTTATTTCACCAACAAC 632

```
Db      584 hrGIUIeThr-----Thrl 589
```

631 TCACTTCTTCAATATTCCACCACTTCTTACGTGACGGATCAGGTCGTTATCATCCCTTC 572

Db 589 eupProglYserThrThrProglYleuSerGluAlaSerThrThrPheThyYserSerP 609

571 CATTAACCAATGAATAACACCACTTCCAGCTGGCGGATATCCGATGGCTGTGGCAGAC 512

Db 609 rohrSerProThrThrIeuSerProIaSerMetThrSerLeuGIValGIgIu- 628

511 TGATCACTGATTCCACTCTGTAAGTTGACCA TTGAGCCAGCATTAACAATTGCAATAT 452

629 -----G|u\$erThrtHr\$erAr0\$ 635

[illegible]

06 635 0701 nDvnc1 wGawthx U4eGawthwiv 644

07 321 CGm1cCn1wNwV1C1W1ScuM1wG1c1Uc1C1W1C1W1CcCnM1nCnMnSbSbC 322

```
QY 391 TGTATCTGGTCCACCTTAATCTCCAGACGGTTGGTGTGATGATGCTTGTGGTCCAT 332
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 644 alserProalaserThrThrThrProgl----- 653
QY 331 TCTTCAGTGTGAAATGAAACCTGAAATTAAGACCAAGTATATAGACATTAAATCTAT- 273
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 654 --LeuSerGluGluSerThrThr-----ValTyrS 663
QY 272 --GGAACCCGCTGATTCACAAACAAGACGACACGCAACACTGCTGTGAAA 215
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 663 erSerSerProGlySerThrGluThr-----ThrVal----- 673
QY 214 TCATTCGATTACATTCAGCAACAATGTGTTTCAGTCGATCTCTTACCGAATCAGTGCA 155
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 674 -----PheProArgSerThrThrThrSerValArg 684
QY 154 ATGTGACCAACA-----CAATGACAGCTGTGATGACTCTCTCGTTGAGACTA 104
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 684 lYglGluProThrThrPheHisSerArgProAlaserThrHisThrThrLeuPheThrg 704
QY 103 GTGAASCTGTACACTGTCTCTCTCTGTCGTTAAGATGCGATGAGCGGAATTCAGTGC 44
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 704 luepSerThrThrSerGlyLeuThrGluGluSerThrAlaPheProGlySerProAlas 724
QY 43 GGTGTTGCACAGGTTCAACCACTACGTATCAACAGGTTGAC 3
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 724 erThrGlnThrGlyLeuProAlaThrLeuThrThrAlaasp 737

RESULT 4
US-09-833-263-1065
: Sequence 1065, Application US/09833263
: Patent No. US20020110547A1
: GENERAL INFORMATION:
: APPLICANT: Wang, Aijun
: APPLICANT: Clapper, Jonathan D.
: APPLICANT: Stolk, John A.
: APPLICANT: Meagher, Madeleine J.
: TITLE OF INVENTION: COMPOUNDS FOR IMMUNOTHERAPY AND
: TITLE OF INVENTION: DIAGNOSIS OF COLON CANCER AND METHODS FOR THEIR USE
: FILE REFERENCE: 210121.471C12
: CURRENT APPLICATION NUMBER: US/09/833,263
: CURRENT FILING DATE: 2001-04-10
: NUMBER OF SEQ ID NOS: 1093
: SOFTWARE: FASTSEQ for Windows Version 3.0
: SEQ ID NO 1065
: LENGTH: 957
: TYPE: PRT
: ORGANISM: Homo sapiens
US-09-833-263-1065

Alignment Scores:
Pred. No.: 0.00225 Length: 957
Score: 121.00 Matches: 118
Percent Similarity: 30.29% Conservative: 38
Best Local Similarity: 22.91% Mismatches: 179
Query Match: 4.54% Indels: 180
DB: 10 Gaps: 24

US-10-020-441-1 (1-1477) x US-09-833-263-1065 (1-957)
QY 1366 ACCTTGAAGGATGGAGAGGACCTGACCTGCA-----CTGAGACGGGTA 1316
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 344 ThrSerSerGlyValSerGluGluSerThrThr--SerHisSerArgProGlySerThrHis 363
QY 1315 CTGTCCGAATTCGTGACCTGCTTTCACACATATAGGATTCGATGGTGCATGTCGGCA 1256
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 363 eThrThrAlaPheProArgSerThrThrThrProGlyLeu-----SerArgHis 379
QY 1255 TTCGATT-----ATAGTCGCTGACACCTGTGTGTC 1226
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 379 eSerThrThrSerHisSerSerProGlySerThrAlaPheThrThrLeuProAlaSerTh 399
```

```
QY 1225 AAACAAACACCAACAACAAGTACAGTACACAACCTGACGTACAGATGCAAGAAACA 1166
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 399 rThrThrSerGlyProSerGlnGluSerThrThrSerHisSerSerProGlySerThrHis 419
QY 1165 AAACAAACGGAACAAACAAACAAACCAAGTGAATTCGAAATATATACATGGCAAT 1106
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 419 pThrAlaLeuSerProGlySerThr-----ThrAlaLeuSerPheGlyGlnG 435
QY 1105 ATCCACCTTTTAC-----ACTACCACACACA----- 1078
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 435 uSerThrThrPheHisSerSerProGlySerThrHisThrThrLeuPheProArgSerTh 455
QY 1077 -----ATAGTCCATATCAATCAACATGTCGACTATATTCAGCATCAAGTCA 1031
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 455 rThrSerSerGlyIleValGlnAlaSerThrArgValHis-----SerSerThrgl 472
QY 1030 CTCTCCATCTCACCACTCTC-----C 1010
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 472 ySerProArg--ThrThrLeuSerProAlaSerSerThrSerProGlyLeuGlnGlyus 492
QY 1009 AAACACATATCAAAACACATTCGACCCCTACTTCATGTACACAACACACTTCATCGTTT 950
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 492 erThrAlaPheGlnThrHisProAlaSerThrHisThrThrProSerThr----- 508
QY 949 GCATGATATTTTCATGAGCTCTTCTCTCCAGTGTCTCAATCATCCATC----- 897
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 509 -----ProSerThrAlaThrAlaProValGluGlnus 519
QY 896 ----ACATCATTTCCATCCCATTCATTCATCCATCACTCACTCAGTCA----- 852
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 519 erThrThrThrHisArgSerProSerSerThrPro-----ThrThrHisPheProAlas 537
QY 851 --CACGTACCTCACCTCACCA-----CAACATGACCGGATGANA 812
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 537 erSerThrThrSerGlyHisSerGlyusSerThrHisPheHisSerSerProAlaHis 556
QY 811 ATCACTGTCTACTGAACTGTACAAAGTCAATCCAAACAGTTGTTATTCAAAGCTC 752
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 557 ----SerGlyThrThrProSerSerAlaHis----- 565
QY 751 TGACCAATACAAACCGGATTCAGTACAGACAGACACAAACAGATACAGACAC 692
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 566 ----SerThrThrSerGlyArgGlyGluSerThrThrSerArgIleSerProGlySerT 584
QY 691 ACAGATACACTGACTGACATCTGACTGACGACTGATGCTGTTATTCCAAACAAC 632
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 584 hrgIuIleThr-----Thrl 589
QY 631 TCATTTCTTCATATTCACACCATTCCTTACGTGACGATCAGCTGTTATCATCCCTTC 572
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 589 euProGlySerThrThrThrProGlyLeuSerGluAlaSerThrThrPheThrSerSerP 609
QY 571 CATTAACCAACATGAAACACCAAGTTCAGGTGGCGGATATCCGATGGCTGTGCAGAC 512
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 609 roArgSerProThrThrThrLeuSerProAlaSerMetThrSerLeuGlyValGlyGlu 628
QY 511 TGATACCTGTGATTCACCTGCTGTAAGTGAACATTGAGCCGACATTCATCAATGAAAT 452
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 629 -----GluSerThrThrSerArgS 635
QY 451 CGAATTCACGTAGTGTCTGTGGATGTTGCTCTCTGTGTCGTCGACATACAAAGGGCA 392
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 635 erGlnProGlySerThr-----HisSerThrv 644
QY 391 TGTATCTGGTCCACCTTAATCTCCAGACGGTTGGTGTGATGATGCTTGTGTGCCAT 332
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 644 alserProalaserThrThrThrProgl----- 653
QY 331 TCTTCAGTGTGAAATGAAACCTGAAATTAAGACCAAGTATATAGACATTAAATCTAT- 273
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 654 --LeuSerGluGluSerThrThr-----ValTyrS 663
QY 272 --GGAACCCGCTGATTCACAAACAAGACGACACGCAACACTGCTGTGAAA 215
```





Db 117 ----- 117  
Qy 442 AGTGAATCGATTGATTTGATGCTGGCTCAATGGTCACTTACAGATGGAATC 501  
Db 118 ThrGlyAlaAspIleAlaLeuLeuGluGluProValAsnIleSerSerArgVal 137  
Qy 502 AGAGTACAGTCTGCCACAGCATCGGATATCCCGCACCTGGAATGGTGTTCAT 561  
Db 138 HistHValMetLeuProAlaSerGluThrPheProGlyMetProCysTrpVal 157  
Qy 562 GTTGTTATGCAAGGAGATGATACGAC 588  
Db 158 ThrGlyTrpGlyAspValAspAsnAsp 166  
RESULT 7  
US-09-923-779-152  
; Sequence 152, Application US/09923779  
; Patent No. US20020076721A1  
; GENERAL INFORMATION:  
; APPLICANT: Pyle, Ruth A.  
; APPLICANT: Xu, Jiangchun  
; APPLICANT: Kalos, Michael D.  
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY  
; TITLE OF INVENTION: AND DIAGNOSIS OF PANCREATIC CANCER  
; FILE REFERENCE: 210121.553  
; CURRENT APPLICATION NUMBER: US/09/923,779  
; CURRENT FILING DATE: 2001-08-06  
; NUMBER OF SEQ ID NOS: 155  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 152  
; LENGTH: 270  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-923-779-152  
Alignment Scores:  
Pred. No.: 0.00291 Length: 270  
Score: 118.00 Matches: 43  
Percent Similarity: 34.33% Conservative: 26  
Best Local Similarity: 21.39% Mismatches: 60  
Query Match: 4.39% Indels: 72  
DB: 10 Gaps: 4  
US-10-020-441-1 (1-1477) x US-09-923-779-152 (1-270)  
Qy 25 GGTGAACCTGTGCACACCGCACT----- 48  
Db 19 GlyProProSerSerHisSerSerArgValHisGlyGluAspAlaValProTyr 38  
Qy 49 GAATCCCGCTCATCGCATTTCTTAACGACAGAGAACCA-----ATGTGT 93  
Db 39 SerTrpProTrpGlnValSerLeuGlnTyrGluTyrSerGlySerPheTyrHisTrpCys 58  
Qy 94 ACAGSTTCACTAGTCTCAACGAGAGAGACTACAGCTGATTTGTTGGCTCACA 153  
Db 59 GlyGlySerLeuIleAlaIleProAspTrpValValThrAlaGlyHisCysIleSerArgAsp 78  
Qy 154 TTGCCAGTATTCGGGTAGAGATGACTGAAACACATTTGGCTCAATGATGATG 213  
Db 79 Leu----- 79  
Qy 214 ATTTCACAGACAGTGTTCGTCGTCGTCGTTGTGTGTGTATGACAGTGGTTGCA 273  
Db 79 ----- 79  
Qy 274 TAGATTTAATGCTATTAATGCTGTTGCTTATTTACAGTTTCACTTCTCACAGAGAA 333  
Db 80 -----ThrTyrGlnValValLeuGlyGluTyrAsnLeuAlaValLysGlu 94  
Qy 334 GCGGACCAACAGGATCCATCATCACCAACCGCTCGAGTTAAAGTGGACACAGATACAG 393  
Db 95 GlyProGluGlnValIleProIleAsnSerGluGluLeuPheValHisProLeuTrpAsn 114

Qy 394 CCTCTTGATATGCGGACGACGACAGAGACCAATCGCACAGACATCGATGATTCAT 453  
Db 115 ArgSerCysValAlaCys-----GlyAsnAsp 123  
Qy 454 ATTGCATTTGTAATGCTGTGCTCAATATGCTCAACTTACAGAGTGAATAGATGATCAGT 513  
Db 124 IleAlaLeuIleLeuLysSerArgSerAlaGlnLeuGlyAspAlaValGlnLeuAlaSer 143  
Qy 514 CTGCCACAGCATCGGATATCCCGCACCTGGAATGGTGTTCATTTGTTATGGA 573  
Db 144 LeuProProAlaGlyAspIleLeuProAsnLysThrProCysTyrIleThrGlyTrpGly 163  
Qy 574 AGG 576  
Db 164 Arg 164  
RESULT 8  
US-09-925-297-695  
; Sequence 695, Application US/09925297  
; Patent No. US20020081659A1  
; GENERAL INFORMATION:  
; APPLICANT: Rosen et al.  
; TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies  
; FILE REFERENCE: PA105  
; CURRENT APPLICATION NUMBER: US/09/925,297  
; CURRENT FILING DATE: 2001-08-10  
; PRIOR APPLICATION NUMBER: PCT/US00/05989  
; PRIOR FILING DATE: 2000-03-08  
; PRIOR APPLICATION NUMBER: 60/124,270  
; PRIOR FILING DATE: 1999-03-12  
; NUMBER OF SEQ ID NOS: 928  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 695  
; LENGTH: 273  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: SITE  
; LOCATION: (127)  
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids  
; NAME/KEY: SITE  
; LOCATION: (28)  
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids  
; NAME/KEY: SITE  
; LOCATION: (34)  
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids  
US-09-925-297-695  
Alignment Scores:  
Pred. No.: 0.00412 Length: 273  
Score: 116.50 Matches: 37  
Percent Similarity: 35.80% Conservative: 21  
Best Local Similarity: 22.84% Mismatches: 49  
Query Match: 4.34% Indels: 55  
DB: 10 Gaps: 2  
US-10-020-441-1 (1-1477) x US-09-925-297-695 (1-273)  
Qy 91 TGTACAGSTTCACTAGTCTCAACGAGAGAGACTACAGCTGTCATTTGTTGCTCA 150  
Db 61 CysGlyGlySerLeuIleAlaIleProAspTrpValValThrAlaGlyHisCysIleSerArg 80  
Qy 151 CCATTGCCAGTATTCGGGTAGAGATGACTGAAACACATTTGGCTCAATGATATGCA 210  
Db 81 AspLeu----- 82  
Qy 211 TTGATTTCACAGACAGTGTTCGTCGTCGTCGTTGTGTGTGTATGACAGTGGTTT 270  
Db 82 ----- 82  
Qy 271 GCATGATTTTAATGCTATTAATGCTGTTGCTTATTTACAGTTTCACTTCTCACATGAGG 330

Db 83 -----ThirtyGlnValValLeuGlyGlnTyrAsnLeuAlaValLys 96  
QY 331 AATGGCAACAAGGATCCATGACCAACCGCTGTGAGTTAAGTGGACAGGATAC 390  
Db 97 GluGlyProGlnValLeuProIleAsnSerGlnGlnLeuPheValHisProLeuTyr 116  
QY 391 ATGCCCTCTTGATGTGGCAGCAGACAGAGAACCAATGACACAGACTGTGGATTTC 450  
Db 117 AsnArgSerCysValAlaLys-----GlyAsn 125  
QY 451 GATATTGCAATGTATGTGGTGCATATGTCATCTTACAGATGATGATCAGATGATC 510  
Db 126 AspIleAlaLeuIleLeuLeuSerArgSerAlaGlnLeuGlyAspAlaValGlnLeuAla 145  
QY 511 AGTCTGCACAGCCATCGGATATCCCGCCACTGGAACTGTGTGTTTCATTTGGTTAT 570  
Db 146 SerLeuProProAlaGlyAspIleLeuProAlaLeuLysThrProCysTyrIleThrGlyTyr 165  
QY 571 GGAAGG 576  
Db 166 GlyArg 167

## RESULT 9

US-10-117-323-35  
; Sequence 35, Application US/10117323  
; Publication No. US20030054993A1  
; GENERAL INFORMATION:  
; APPLICANT: Rancourt, Derrick E.  
; APPLICANT: Rancourt, Susan L.  
; APPLICANT: O'Sullivan, Colleen M.  
; TITLE OF INVENTION: Implantation Serine Proteinases  
; FILE REFERENCE: 033337-005  
; CURRENT FILING DATE: 2002-04-08  
; PRIOR FILING DATE: 2001-04-06  
; PRIOR APPLICATION NUMBER: US 60/281,724  
; PRIOR FILING DATE: 2001-04-06  
; PRIOR APPLICATION NUMBER: US 60/294,736  
; PRIOR FILING DATE: 2001-05-30  
; PRIOR APPLICATION NUMBER: US 60/350,962  
; PRIOR FILING DATE: 2002-01-25  
; NUMBER OF SEQ ID NOS: 41  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 35  
; LENGTH: 276  
; TYPE: PRT  
; ORGANISM: Mouse  
US-10-117-323-35

## Alignment Scores:

Pred. No.: 0.0147 Length: 276  
Score: 111.00 Matches: 41  
Percent Similarity: 34.86% Conservative: 20  
Best Local Similarity: 23.43% Mismatches: 38  
Query Match: 4.13% Indels: 76  
DB: 9 Gaps: 5

US-10-020-441-1 (1-1477) x US-10-117-323-35 (1-276)

QY 91 TGTACAGTTACTAGTGTCTCAACAGACAGTACTCAGCTGTGATTTGTT----- 144  
Db 60 CysGlyGlySerLeuIleHisProGlnTyrValLeuThrAlaIleHisCysValAlaGlyPro 79  
QY 145 -----TGTCAACATTCGACGTATGCGGTTAGAGATTCAGTGAACACATT----- 192  
Db 80 HisIleLeuSerPro--GlnLeuPheArgValGlnLeuArgGlnGlnIleuTyr 98  
QY 193 -----GTTGTCATATGATTCAGATTTTCACAGACAGTGTTCGCTGTGCTGC 243  
Db 99 GlyAspGlnLeuLeuSerLeuAsnArgIleVal----- 109  
QY 244 GTTGTGTGTTGTGATGACAGTGGGTTTGCATGATTTTAATGCTTAATCTGCTTTA 303  
Db 109 ----- 109

QY 304 TTTCAGTTTCAATTCTCACACTGAGGAATGGCGCAACAAGGATTCATCACAACCG 363  
Db 109 ----- 109  
QY 364 TGTGAGTTAAGTGGACACAGATACATGCCCTTGTGATGTGGCAGACAGAGAGA 423  
Db 110 -----ValHisProHisTyrTyrThrAla----- 117  
QY 424 CCAATGCAACAGACTCAGTGGATTCGATATGTGCAATTTGATATGCTGCTCAATGTC 483  
Db 118 -----GlnGlyGlyAlaAspValAlaLeuLeuGlnLeuGlyValProVal 132  
QY 484 AACTTACAGAGTGAATGACAGTGTAGTCACTGCGCACAGCCATCGGATATCCCGCCACT 543  
Db 123 AsnValSerThrHisIleHisProIleSerLeuProProAlaSerGlnThrPheProPro 152  
QY 544 GGAAGTGTGTTTCAATGTTGTTGTTATGAGAGGATGATACAGAC 588  
Db 153 GlyThrSerCysTyrPValThrGlyTyrPglYAspIleAspAsnAsp 167

## RESULT 10

US-10-117-323-40  
; Sequence 40, Application US/10117323  
; Publication No. US20030054993A1  
; GENERAL INFORMATION:  
; APPLICANT: Rancourt, Derrick E.  
; APPLICANT: Rancourt, Susan L.  
; APPLICANT: O'Sullivan, Colleen M.  
; TITLE OF INVENTION: Implantation Serine Proteinases  
; FILE REFERENCE: 033337-005  
; CURRENT FILING DATE: 2002-04-08  
; PRIOR FILING DATE: 2001-04-06  
; PRIOR APPLICATION NUMBER: US 60/281,724  
; PRIOR FILING DATE: 2001-04-06  
; PRIOR APPLICATION NUMBER: US 60/294,736  
; PRIOR FILING DATE: 2001-05-30  
; PRIOR APPLICATION NUMBER: US 60/350,962  
; PRIOR FILING DATE: 2002-01-25  
; NUMBER OF SEQ ID NOS: 41  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 40  
; LENGTH: 272  
; TYPE: PRT  
; ORGANISM: Mouse  
US-10-117-323-40

## Alignment Scores:

Pred. No.: 0.0259 Length: 272  
Score: 108.50 Matches: 36  
Percent Similarity: 35.12% Conservative: 23  
Best Local Similarity: 21.43% Mismatches: 46  
Query Match: 4.04% Indels: 63  
DB: 9 Gaps: 3

US-10-020-441-1 (1-1477) x US-10-117-323-40 (1-272)

QY 91 TGTACAGTTACTAGTGTCTCAACAGACAGTACTCAGCTGTGATTTGTTGCTCA 150  
Db 57 CysGlyGlySerLeuIleHisProGlnTyrValLeuThrAlaIleHisCysValAlaGlyPro 76  
QY 151 CCAATGCAACAGATTCG-----GTAAAGATGAGACTGGAACACATTTGCTCAATGT 204  
Db 77 AspValIlyAspLeuAlaAlaLeuValGlnLeuArgGlnGlnHisLeuTyrTyrGlnAsp 96  
QY 205 AATGATGATTTTCACAGACAGTGTTCGCTGTGCGTGTGTGTGTGTGTGATGACAGT 264  
Db 97 Gln----- 97  
QY 265 GGGTTTGATGATTTTAATGCTTAATCTGCTTTATTTCAAGTTTCAATTTCTACA 324  
Db 98 -----LeuLeuProValSerArgIleIleValHisProGlnPheTyrThr 112

QY 325 CTGAGGAATGGCGACCAAGGATCATCAACACCGTGTGAGTTAAGTGGCACCA 384  
Db 113 AlaGln----- 114  
QY 385 GGATACATGCCCTCTTGTATGTGCGACGACAGAGAGACCAATGCGACAGACTCACT 444  
Db 115 -----11e 115  
QY 445 GGATTCGATATTCGCAATGTAATGCTGGCTCAATGTGTCATCTTACAGATGGAATCAGA 504  
Db 116 GtYlaaApIleAlaLeuLeuGluLeuGluProValAsnValSerIstValHis 135  
QY 505 GTGATCACTGTCCACAGCCATCGGATATCCCGCACCTGGAACTGGTGTTCATTGTT 564  
Db 136 ThrValThrLeuProAlaSerGluThrPheProGlyMetProCysIstPValThr 155  
QY 565 GGTATGGAAGGATGATACAC 588  
Db 156 GlyTrpGlyAspValAspAsnAsp 163

## RESULT 11

US-10-083-357-1014  
; Sequence 1014, Application US/10083357  
; Publication No. US20030054370A1  
; GENERAL INFORMATION:  
; APPLICANT: Qidong Zeng et al.  
; TITLE OF INVENTION: Systemic Discovery of New Genes  
; FILE REFERENCE: 032796-090  
; CURRENT APPLICATION NUMBER: US/10/083.357  
; CURRENT FILING DATE: 2002-02-27  
; NUMBER OF SEQ ID NOS: 1346  
; SEQ ID NO 1014  
; LENGTH: 153  
; TYPE: PRT  
; ORGANISM: Saccharomyces cerevisiae  
US-10-083-357-1014

Alignment Scores:  
Pred. No.: 0.0268 Length: 153  
Score: 107.50 Matches: 45  
Percent Similarity: 44.97% Conservative: 22  
Best Local Similarity: 30.20% Mismatches: 38  
Query Match: 4.03% Indels: 44  
Gaps: 9

US-10-020-441-1 (1-1477) x US-10-083-357-1014 (1-153)

QY 1111 CAGAAATATCCACCT--TTAACAATCCACACAAACATATGTCATATACATATGT 1055  
Db 9 HistHrProHIsrProHIsrThrHIsrProHIsrThrProHIsrThr-Pr 28  
QY 1054 GCACATATATTCAGCATCAAGTCACTGTCATCTGACCTGCCAAACAAACATATGCA 995  
Db 28 ohIs-----ProHIsrThrProHIsrThrHIsrThrProHIsrThrHIsrThrH 43  
QY 994 CACATTCAGCCCCCTATCTTCATGATACACAAACATCTTCATGTCATATATTCAT 935  
Db 43 sHrProHrProHIsrProHIsrThrHIsrThr----- 55  
QY 934 GAGTCTTCTCTCCCAAGTGTCTCAATCATCCATCAACATCACT----- 888  
Db 56 -----ProHrProHrProHIsrProHIsrThrHIsrThrProHrThrL 71  
QY 887 --TTCATCCCA--TTCATTCATCCCACTGACCTGACATGACGTCACCTGACCTGAC 833  
Db 71 eutHrLeuProTySerAsnProAspGluProValSerGln-LeuThrProProLeuPro 90  
QY 832 AACACATGACCGAGATGGAATCACTCTGTACTGATCACTGTCACAAATC-----ATC 779  
Db 91 AnSerSerPhe-----ValThrLeu-----CysHIsrThrLeuProIle 104  
QY 778 CAACACAGTTGTTATATATCAAGGCTCTGACCAATACAAACCGGATCAGGTAGACAGA 719

Db 105 ProHrThrIleHIsrHIsrAsnArgTyProProIle----- 116  
QY 718 CAGACACACAAACAGATACAC 698  
Db 117 ---ThrHIsrIleGlnLeuHIs 122

## RESULT 12

US-10-025-380-1068  
; Sequence 1068, Application US/10025380  
; Publication No. US20020182191A1  
; GENERAL INFORMATION:  
; APPLICANT: Xu, Jianshun  
; APPLICANT: Lodes, Michael J.  
; APPLICANT: Secret, Heather  
; APPLICANT: Benson, Darin R.  
; APPLICANT: Meagher, Madeleine Joy  
; APPLICANT: Stolk, John A.  
; APPLICANT: Wang, Tonglong  
; APPLICANT: Jiang, Yugu  
; APPLICANT: Smith, Carole L.  
; APPLICANT: King, Gordon E.  
; APPLICANT: Wang, Aijun  
; APPLICANT: Clapper, Jonathan D.  
; APPLICANT: Skeiky, Yasir A. W.  
; APPLICANT: Fanger, Gary R.  
; APPLICANT: Vedrick, Thomas S.  
; APPLICANT: Carter, Darick  
; TITLE OF INVENTION: COMPOUNDS FOR IMMUNOTHERAPY AND DIAGNOSIS  
; FILE REFERENCE: 210121.471C14  
; CURRENT APPLICATION NUMBER: US/10/025.380  
; CURRENT FILING DATE: 2001-12-19  
; NUMBER OF SEQ ID NOS: 1129  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 1068  
; LENGTH: 5179  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-10-025-380-1068

Alignment Scores:  
Pred. No.: 0.178 Length: 5179  
Score: 104.50 Matches: 84  
Percent Similarity: 30.33% Conservative: 27  
Best Local Similarity: 22.95% Mismatches: 169  
Query Match: 3.92% Indels: 86  
Gaps: 12

US-10-020-441-1 (1-1477) x US-10-025-380-1068 (1-5179)

QY 1392 CCATGTGATACGACACACGAGTACTGACCTTGAAGGATGGAGGAGGATCCACACTG 1333  
Db 1557 ProProHrThrThrProSerProProHrThrThrThrThrThrThrThrThrThr 1576  
QY 1332 TCACCTGGAGCGGTAACTGTCCGAAATTCGACCTGCTTACACATATAGATTGCCA 1273  
Db 1577 ThrProSerPro-----ProHrThrThrThrThrThrThrThrThrThrThr 1589  
QY 1272 TTGGTCGATGTCGGCATTCATATATAGTCGTGACCCCTGTGGCAAAACAAACACA 1213  
Db 1590 ThrIleThrThr-----ThrThrProProPro 1598  
QY 1212 CAACAAAGTACGATACAAACAGTACGATACGATGAGAAACAAACAAACGAA 1153  
Db 1599 ThrThrThrProSerProProHrThrThrThrThrThrThrThrThrThrThrThr 1618  
QY 1152 CAAAACAAACACACAAAGTAAATTTCG---AAAATATCAACTGGCAATATTCACCTT 1096  
Db 1619 SerProProHrThrThrThrProIleHrProProHrThrThrThrThrThrThrThr 1638  
QY 1095 TACACTACCCACACAAACATATAGTCCACATACATACATGTCACATATTTGACAGATA 1036  
Db 1638 rThrThrProSerProProProHrThr----- 1647

```

QY 1035 AGTCACTCTGCATCTACCACTCTCCAAACACATCTCAACACATTTGACCCCTACTTC 976
Db 1648 -----ThrThrThrProProThrThrThrThrProSerProThrThr 1662
QY 975 ATGTACACACACACTTCATCTGTTGCATAGATATTTCAATAGTCTCTCTCCAG 916
Db 1662 rThrThrProSerProProThrThrThrThrThrThrProProThrThrProse 1682
QY 915 TGGTCTCAATATCCCATCATCATTCATCCATCCCATCATCTCACTCACTCC 856
Db 1682 rSer-----ProThrThrThrThrProSerProProThrThrThrMetThrThrPr 1699
QY 855 ACATCAGCTCACTCACTCACTCACTCACTCACTCACTCACTCACTCACTCACTCA 796
Db 1699 oSerProThrThrThrThrProSerSerProThrThrThrThrThrThrProSerThrThr 1719
QY 795 ACTGTCAAAAGTCAATCCAAACAGTTGTTATTCAGAGCTTGACCAATACAAAC 736
Db 1719 rThrProSerProProProThrThrThr-----ThrProSerProThrThrThr 1736
QY 735 CGGATCAGTACAGACAGACACACAAACATACAGACACACAGAAATACAGCTG 676
Db 1736 rProSerProProThrThrThrMetThrThrLeuProProThr----- 1750
QY 675 ACATCTGACGACGACGATGATGATGCTTTATTCACCAACACTCACTTCTTCATAT 616
Db 1751 -----ThrThrSerSerProLeuThrThrThrProle 1761
QY 615 TCACACATTTTACGTACGATCAGATCAGGTCGTTATTCATCCCTTCATACCAACATGA 556
Db 1761 uProSerThrThrThrProThrThrProThrProThrProThr----- 1779
QY 555 AACCCAGTT----- 546
Db 1780 -ThrProCysValProLeuCysAsnTyrThrGlyTyrLeuAspSerGlyLysProAsnPh 1799
QY 545 -----CCAGTGGCGGG-----ATATCCGATGCTGTGGCAGACTGATCACTCT 502
Db 1799 eHisLysProGlyGlyAspThrGlyLeuLeuLeuGlyAspValCysGlyProGlyTyrPalaAl 1819
QY 501 GATTCCACTCTGTAAGTTCAGCATTTGAGCCAGCATTCATCAATTCGAATTCACCT 442
Db 1819 AsnLysSerCysArgAlaThrMetLysProAspValProThrGlyGlyLeuGlyGlyThr 1839
QY 441 GAGTGTCTGT-----GGATTGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 388
Db 1839 rValValCysAspValSerValGlyLeuLeuLeuCysLysAsnGlyAspGlnLys----- 1856
QY 387 TCCTGTCGCACCTTA 372
Db 1857 -ProGlyGlyValIle 1861

RESULT 13
US-09-922-217-1068
; Sequence 1068, Application US/09922217
; Patent No. US20020076414A1
; GENERAL INFORMATION:
; APPLICANT: Xu, Jiangchun
; APPLICANT: Lodes, Michael J.
; APPLICANT: Secretat, Heather
; APPLICANT: Benson, Darin R.
; APPLICANT: Mesgher, Madeleine Joy
; APPLICANT: Stoik, John A.
; APPLICANT: Wang, Tonglong
; APPLICANT: Jiang, Yugu
; APPLICANT: Smith, Carole Lynn
; APPLICANT: King, Gordon E.
; APPLICANT: Wang, Aijun
; APPLICANT: Clapper, Jonathan D.
; TITLE OF INVENTION: COMPOUNDS FOR IMMUNOTHERAPY AND DIAGNOSIS
; FILE REFERENCE: 210121.471C13

```

```

; CURRENT APPLICATION NUMBER: US/09/922.217
; CURRENT FILING DATE: 2001-08-03
; NUMBER OF SEQ ID NOS: 1124
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1068
; LENGTH: 5179
; TYPE: PR
; ORGANISM: Homo sapiens
US-09-922-217-1068

Alignment Scores:
Pred. No.: 0.178 Length: 5179
Score: 104.50 Matches: 84
Percent Similarity: 30.33% Conservative: 27
Best Local Similarity: 22.95% Mismatches: 169
Query Match: 3.92% Indels: 86
Db: 10 Gaps: 12

US-10-020-441-1 (1-1477) x US-09-922-217-1068 (1-5179)
QY 1392 CCATGTGATAGACACGAGTACTGACCTTGAAGGATGGAGAGAGTCCACCACTG 1333
Db 1557 ProProThrThrThrThrProSerProProProThrThrThrThrProProThrThr 1576
QY 1332 TCACCTGAGGCGGATTCGTAATTCGAAATTCGACCTGCTTCACACATATGATTCGA 1273
Db 1577 ThrProSerPro-----ProThrThrThrThrProSerProPro 1589
QY 1272 TTGTCGATGTCGATTCATTAATAGTCGCTGACCTGTGTGCAACAAACACCA 1213
Db 1590 ThrThrThr-----ThrThrProProPro 1598
QY 1212 CAACACAGTACAGTACACACGATACAGTACAGTACAGTACAGTACAGTACAGTACAG 1153
Db 1599 ThrThrThrProSerProProThrThrThrThrThrThrThrProProProThrThrPro 1618
QY 1152 CAACACAGTACAGTACACGATTCG-----AAATATCACTGCGACAAATATCCACCTT 1096
Db 1619 SerProProThrThrThrProThrProThrProThrProThrProThrProThrProThr 1638
QY 1095 TACACTCCACACACAAATAGTCCACATACATACATACATACATACATACATACATAC 1036
Db 1638 rThrThrProSerProProProThrThr----- 1647
QY 1035 AGTCACTCTGCATCTCACTCTCCAAACACATCTCAACACATTTGACCCCTACTTC 976
Db 1648 -----ThrThrThrProProProThrThrThrProSerProProThrThr 1662
QY 975 ATGTACACACACACTTCATCTGTTGCATAGATATTTCAATAGTCTCTCTCCAG 916
Db 1662 rThrThrProSerProProThrThrThrThrThrThrProProProThrThrThrProse 1682
QY 915 TGGTCTCAATATCCCATCATCATTCATCCATCCCATCATCTCACTCACTCC 856
Db 1682 rSer-----ProThrThrThrThrProSerProProThrThrThrMetThrThrPr 1699
QY 855 ACATCAGCTCACTCACTCACTCACTCACTCACTCACTCACTCACTCACTCACTCA 796
Db 1699 oSerProThrThrThrThrProSerSerProThrThrThrThrThrThrProSerThrThr 1719
QY 795 ACTGTCAAAAGTCAATCCAAACAGTTGTTATTCAGAGCTTGACCAATACAAAC 736
Db 1719 rThrProSerProProProThrThrThr-----ThrProSerProThrThrThr 1736
QY 735 CGGATCAGTACAGACACACACAAACAGATACACAGACACAGAAATACAGCTG 676
Db 1736 rProSerProProThrThrThrMetThrThrLeuProProThr----- 1750
QY 675 ACATCTGACGACGACGATGATGCTTTATTCACCAACACTCACTTCTTCATAT 616
Db 1751 -----ThrThrSerSerProLeuThrThrThrProle 1761
QY 615 TCACACATTTTACGTACGATCAGATCAGGTCGTTATTCATCCCTTCATACCAACATGA 556

```

```
Db 1761 uProProSerIleThrProProThrPhSerProPhSerThrThrProThr----- 1779
Qy 555 AACACCACTT----- 546
Db 1780 -ThrProCysValProLeuCysAsnTrpThrGlyTrpLeuAspSerGlyLysProAsnPh 1799
Qy 545 -----CCAGGTGGCGGG-----ATATCCGATGGCTGTGGCAGACTGATCACTCT 502
Db 1799 eHisLysProGlyLysAspThrGluLeuIleGlyAspValCysGlyProGlyTrpAla1 1819
Qy 501 GATTCACCTGTGAAGTTGACCATTTGAGCCAGCATTTACATTCGCAATATCGAATCCACT 442
Db 1819 aAsnLysSerCysArgAlaThrMetLysProAspValProIleGlyGlnLeuGlyGlnTh 1839
Qy 441 GAGTGTCTGT-----GGGATTGGTCTCTCTGTGTGGCCGACATACAAAGGGCATGTA 388
Db 1839 rValValCysAspValSerValGlyLeuIleCysLysAsnGlnAspGlnLys----- 1856
Qy 387 TCCTGTGGCCACCTTA 372
Db 1857 -ProGlyGlyValIle 1861
```

## RESULT 14

US-09-833-263-1068  
; Sequence 1068, Application US/09833263  
; Patent No. US20020110547A1

## GENERAL INFORMATION:

; APPLICANT: Wang, Aljun  
; APPLICANT: Clapper, Jonathan D.  
; APPLICANT: Stolck, John A.  
; APPLICANT: Meagher, Madeleine J.  
; TITLE OF INVENTION: COMPOUNDS FOR IMMUNOTHERAPY AND  
; FILE REFERENCE: 210121.471C12  
; CURRENT APPLICATION NUMBER: US/09/833,263  
; CURRENT FILING DATE: 2001-04-10  
; NUMBER OF SEQ ID NOS: 1093  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 1068  
; LENGTH: 5179  
; TYPE: PRT  
; ORGANISM: Homo sapiens

US-09-833-263-1068

## Alignment Scores:

Pred. No.: 0.178 Length: 5179  
Score: 104.50 Matches: 84  
Percent Similarity: 30.33% Conservative: 27  
Best Local Similarity: 22.95% Mismatches: 169  
Query Match: 3.92% Indels: 86  
DB: 10 Gaps: 12

US-10-020-441-1 (1-1477) x US-09-833-263-1068 (1-5179)

```
Qy 1392 CCATGTGATACGACACCGAGTCTGAGCCTTGAAGGATGGAGAGAGGTCACACCACTG 1333
Db 1557 ProProThrThrProSerProProThrThrThrThrProProProThrThr 1576
Qy 1332 TCACCTGAGCGGGTACGTGCGAAATTCGACCTTTCACACATATAGATTGCCA 1273
Db 1577 ThrProSerPro-----ProThrThrThrThrProSerProPro 1589
Qy 1272 TTGGTGGCATGTGGCATTCATATAGTGTCTGCACCCGTGGTGCACAAACAAACCA 1213
Db 1590 ThrIleThrThr-----ThrThrProProPro 1598
Qy 1212 CAACAACAGTACAGTACAAACAGTACAGTACAGTACAGTACAGTACAGTACAGTAC 1153
Db 1599 ThrThrThrProSerProProThrThrThrThrThrThrThrThrThrThrThrPro 1618
Qy 1152 CAAAACAAACACACAAAGTGAATTTCG-----AAAATATCACTGGCACAATATCCACCTTT 1096
```

```
Db 1619 SerProProThrThrThrProIleThrProProThrSerThrThrThrLeu-ProProTh 1638
Qy 1095 TACACTACCCACACAAACATATGTCACATACATACATACATACATATTCGACGATCA 1036
Db 1638 rThrThrProSerProProProThrThr----- 1647
Qy 1035 AGTCACTCTCCATCTGACCACTCTCCAAACACATACATCAAAACACATTACCCCTACTTC 976
Db 1648 -----ThrThrThrProProProThrThrThrThrProSerProProThrTh 1662
Qy 975 ATGTACACACACACTTCATCCATCGTTTGATAGATATTCATGAGTCTCTGCTCCCAAG 916
Db 1662 rThrThrProSerProProIleThrThrThrThrProProProThrThrThrProse 1682
Qy 915 TGGTCTCAATCATCCATACATACATTCATCCATCCATCCATCAATCACTCACTCACTC 856
Db 1682 rSer-----ProIleThrThrThrProSerProProThrThrThrMetThrThrPr 1699
Qy 855 ACATCAGCTACCTCACTCACTCACTCACTCACTCACTCACTCACTCACTCACTCACTCA 796
Db 1699 oSerProThrThrThrProSerSerProIleThrThrThrThrThrProSerSerThrTh 1719
Qy 795 ACTGTACAAAGTCAATCCAAACACAGTTGTTATTATCAAGGCTGCACCAATACAAACAC 736
Db 1719 rThrProSerProProProThrThrThrThr-----ThrProSerProThrThrTh 1736
Qy 735 CGGATCAGTACAGTACACACACACACAAACAGATACAGACACACACAGAAATACAGACTG 676
Db 1736 rProSerProProThrThrThrThrMetThrThrLeuProProThr----- 1750
Qy 675 ACATCTGACGACGACGACGATGATGCTGTTATTATCCAAACACATCACTCTTTCTTAATAT 616
Db 1751 -----ThrThrSerSerProLeuThrThrThrProLe 1761
Qy 615 TCCACCATTTCTTACGTGACGATCAAGGTCGTTATCATCCTTCATCAACAAACATGAA 556
Db 1761 uProProSerIleThrProProThrPhSerProPhSerThrThrThrProThr----- 1779
Qy 555 AACACCACTT----- 546
Db 1780 -ThrProCysValProLeuCysAsnTrpThrGlyTrpLeuAspSerGlyLysProAsnPh 1799
Qy 545 -----CCAGGTGGCGGG-----ATATCCGATGGCTGTGGCAGACTGATCACTCT 502
Db 1799 eHisLysProGlyLysAspThrGluLeuIleGlyAspValCysGlyProGlyTrpAla1 1819
Qy 501 GATTCACCTGTGAAGTTGACCATTTGAGCCAGCATTTACATTCGCAATATCGAATCCACT 442
Db 1819 aAsnLysSerCysArgAlaThrMetLysProAspValProIleGlyGlnLeuGlyGlnTh 1839
Qy 441 GAGTGTCTGT-----GGGATTGGTCTCTCTGTGTGGCCGACATACAAAGGGCATGTA 388
Db 1839 rValValCysAspValSerValGlyLeuIleCysLysAsnGlnAspGlnLys----- 1856
Qy 387 TCCTGTGGCCACCTTA 372
Db 1857 -ProGlyGlyValIle 1861
```

## RESULT 15

US-09-801-368-106  
; Sequence 106, Application US/09801368  
; Patent No. US20020128250A1

## GENERAL INFORMATION:

; APPLICANT: Busby, Robert  
; APPLICANT: Cali, Brian  
; APPLICANT: Hecht, Peter  
; APPLICANT: Holtzman, Doug  
; APPLICANT: Madden, Kevin  
; APPLICANT: Maxon, Mary  
; APPLICANT: Milne, Todd  
; APPLICANT: No. US20020128250A1man, Thea  
; APPLICANT: Royer, John  
; APPLICANT: Salama, Sofie

```

; APPLICANT: Sherman, Amir
; APPLICANT: Silva, Jeff
; APPLICANT: Summers, Eric
; TITLE OF INVENTION: Methods for Improving Secondary Metabolite Production in Fungi
; FILE REFERENCE: 109272.147
; CURRENT APPLICATION NUMBER: US/09/801,368
; CURRENT FILING DATE: 2001-03-07
; PRIOR APPLICATION NUMBER: US 09/487,558
; PRIOR FILING DATE: 2000-01-19
; PRIOR APPLICATION NUMBER: US 60/160,587
; PRIOR FILING DATE: 1999-10-20
; NUMBER OF SEQ ID NOS: 440
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 106
; LENGTH: 1169
; TYPE: prt
; ORGANISM: Saccharomyces cerevisiae
US-09-801-368-106

```

```

Alignment Scores:
Pred. No.: 0.151 Length: 1169
Score: 103.00 Matches: 84
Percent Similarity: 40.61% Conservative: 50
Best Local Similarity: 25.45% Mismatches: 126
Query Match: 3.86% Indels: 70
DB: 10 Gaps: 16

```

US-10-020-441-1 (1-1477) x US-09-801-368-106 (1-1169)

```

QY 1339 ACCACTGTCACCTGAGCGGGTAAGTCTCGAATTCGAC--TGCTTTCACACATAT 1283
    ||||| :||| ||||| ||||| ||||| |||||
DB 544 ThrsrSsrSsrTrpSsrSsr-----SerGluValCythrGluCythrGluThr--- 560
QY 1282 AGGATTGCCATTGGTCGATGTCGCGCATC-----CATTATAGTGGCTCGACCTG 1232
    ||||| :||| ||||| ||||| ||||| |||||
DB 561 -----GluSerThrSsrYrValThrProYrValSsrSsrSsrThr-Al 575
QY 1231 TGGTCAAAACAACAACAACAACAACAACAACAACAACAAGTACAGTACAGATGAG-- 1174
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 575 aalalalalal-----TyrThrSsrSsrPheSsrSsrSsrSsrSsrGluVal 589
QY 1173 ---AAAGACAACAACAACAACAACAACAACAACAACAAGTAAATTCGAAAAATATC 1118
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 589 lcythrGluCythrGluThrGluThrSsrThrSsrThrSsrThrProYrAlaThrSsrSe 609
QY 1117 AACTGGCAAAATATCCACCTTACATCACTACACACAAATATGTCACATATCACA 1058
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 609 ThrGluYrThrAlaThrSsr-PheThrAla-----SerThrSsrAsnThrM 624
QY 1057 TGTGCACTATTCGACGATCAAGTCAATCTTCATCTC-----ACCACCTCCAAA 1007
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 624 etThrSerLeuValGlnThrAlaThrThrValSerPheSsrLeuSsrThrThrValSerg 644
QY 1006 CACATATCTAAACACATTCAGCCCTTACTTCTGTACACACACACTTCATCGTTTGA 947
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 644 luhlaThrAsnAla-----ProThrSsrSerValGluSsrAsnAlaSerThrPheI 661
QY 946 TAGATATTCATGAGTCTTCTCTCTCCCAAGTGTCTCATATATCCCATCATCA----- 891
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 661 lSer-----SerAsnYsglySerValYsglySerYrValThrSsrSeri 676
QY 890 --CATTTCATC-----CCATCCATCTACCTCACTCACTCACTCACTCACTCACTCA 845
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 676 lehlaSerlleThrProMetYrTrpSsrPheAsnGlnThrValThrSsrSsrValValS 696
QY 844 CCTCACTCACCACACACATGACCGAGATGGAATCACTCTCTACTGTAAGT----- 792
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 696 erThrProIleThrSsrGluSsrSsrGluSsrSsrAlaSerValThrIleLeuProSsrT 716
QY 791 -----TCACAAAGTATCCAAACACAGTTGTTATTATCAAGGCTCGACCAATACAA 740
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 716 hrIleThrSsrGluThrSsrProSsrThrMet-----LysThrIlySsrValV 731

```

```

QY 739 CAACCGATCAGTAGACAGACACACACAAACAAGATACACAGACACACATTAACCTG 680
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 731 alSerlleSsrSsrSsrProThrAsnLeuIleThrSsrYrAspThr-ThrSsrIlyAsp 750
QY 679 ACTGACATCTGACTACAGCATGATGATGCTTTATTCACCAACACTCTTCTTCA 620
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 751 SerThrValGlySsrSsrThrSsrSsrValSerlleSsrSsrSsrIleSsrLeuProSsr 770
QY 619 -----ATATTCACCATCTCTTACGTCAGGATCAACGATCGATCGAT 584
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 771 SerYrSsrAlaSerSerGluGlnIlePheIleSsrSerlleValSerSsrAsnIlyGln 790
QY 583 TATCATCCCTTCATACCAACAATGAAAACAAGTTCACAGTGCAGGATATTCGATG 524
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 791 -AlaLeuThrSsrPheSsrSsrThrIlyValSsrSsrSsrGluSsrSsrIlyAsn 810
QY 523 GGTGGGACAGTATCATCTGATTCACCTGTGAATGACATTGAGCCACATTA 464
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 810 gThrSsrProThrThrSsrSsrGluSsrIlyle-----LysSsrSe 824
QY 463 CAATTGCAATATCGAATCCACT 442
DB 824 rGlyValGlnIleGluSsrThr 831

```

RESULT 16  
US-10-012-896-479

Sequence 479, Application US/10012896  
Publication No. US20020183251A1

GENERAL INFORMATION:

```

; APPLICANT: Xu, Jianshun
; APPLICANT: Dillon, Devin C.
; APPLICANT: Mitcham, Jennifer L.
; APPLICANT: Harlocker, Susan L.
; APPLICANT: Jiang, Yugu
; APPLICANT: Kalos, Michael D.
; APPLICANT: Retter, Marc W.
; APPLICANT: Stolk, John A.
; APPLICANT: Day, Craig H.
; APPLICANT: Vedvick, Thomas S.
; APPLICANT: Carter, Derrick
; APPLICANT: Li, Samuel X.
; APPLICANT: Wang, Aljun
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Hepher, William T.
; APPLICANT: Henderson, Robert A.
; APPLICANT: Hurai, John
; APPLICANT: McNeill, Patricia D.
; APPLICANT: Houghton, Raymond L.
; APPLICANT: Vinals de Bassols, Carlota
; APPLICANT: Foy, Teresa
; APPLICANT: Fanger, Gary R.
; APPLICANT: Watanabe, Yoshihiro
; APPLICANT: Meagher, Madeleine Joy
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
; FILE REFERENCE: 210121.427C27
; CURRENT APPLICATION NUMBER: US/10/012,896
; NUMBER OF SEQ ID NOS: 1011
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 479
; LENGTH: 222
; TYPE: prt
; ORGANISM: Homo sapiens
US-10-012-896-479

```

```

Alignment Scores:
Pred. No.: 0.0961 Length: 222
Score: 102.50 Matches: 60
Percent Similarity: 30.95% Conservative: 18
Best Local Similarity: 23.81% Mismatches: 63
Query Match: 3.84% Indels: 111
DB: 9 Gaps: 14

```





```

Db 123 sptnglnasnthValthrgarghnsishislaasprthProphronisgluhs 143
QY 826 ATGACCGAGATGGAATGACTGCTAGTACGACTGCAAGATGATCCAAAGATGAG 767
Db 143 ertharg-----Hisserlatatnglile-----Hisishisthrglum 157
QY 766 TTATATGACAGGCTCTGACCAATACACACACC 735
Db 157 etargthhisCysenishthraerththtr 167

RESULT 18
US-09-895-814-479
; Sequence 479, Application US/09895814
; Publication No. US20020193296A1
; GENERAL INFORMATION:
; APPLICANT: Xu, Jiangchun
; APPLICANT: Dillon, Davin C.
; APPLICANT: Mitcham, Jennifer L.
; APPLICANT: Harlocker, Susan L.
; APPLICANT: Jiang, Yugu
; APPLICANT: Kalos, Michael D.
; APPLICANT: Retter, Marc W.
; APPLICANT: Stolck, John A.
; APPLICANT: Day, Craig H.
; APPLICANT: Vedvick, Thomas S.
; APPLICANT: Carter, Derrick
; APPLICANT: Li, Samuel X.
; APPLICANT: Wang, Aijun
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Hepler, William T.
; APPLICANT: Henderson, Robert A.
; APPLICANT: Houghon, Raymond L.
; APPLICANT: Vinals de Bassols, Carlota
; APPLICANT: Foy, Teresa
; APPLICANT: Fanger, Gary R.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
; TITLE OF INVENTION: DIAGNOSIS OF PROSTATE CANCER
; FILE REFERENCE: 210121.427C26
; CURRENT APPLICATION NUMBER: US/09/895, 814
; NUMBER OF SEQ ID NOS: 990
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 479
; LENGTH: 222
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-895-814-479

Alignment Scores:
Pred. No.: 0.0961 Length: 222
Score: 102.50 Matches: 60
Percent Similarity: 30.95% Conservative: 18
Best Local Similarity: 23.81% Mismatches: 63
Query Match: 3.84% Indels: 111
DB: 9 Gaps: 14

US-10-020-441-1 (1-1477) x US-09-895-814-479 (1-222)
QY 1453 CAACATTCAGACATGAGTACGACATGATATCGGAGAGTTGGAGATGTGAC 1394
Db 16 Glnserhnsghunishthglyleval-----Thtrtp---Thrasp 28
QY 1393 ACCATGTGATGACGAC-----ACCGAGTACGAGCTTGAGGAGATGGAGAG 1346
Db 29 Thrglnthrgylgluilethreleuthrhishishisthilethrglythrglnthr 48
QY 1345 AGGTCCACCACTGTGACCTGAGCGGGTAATCTGCCAAATTCGACCTGCTTTACACA 1286
Db 49 HnsglAerlIerthtrthtr----- 56

```

```

QY 1285 TATAGATTCGCGATTCGTGCGATGTGGCATTCATTATAGTGGCTGACCTGTGGTC 1226
Db 57 -----HisCys 58
QY 1225 AAACAAGACACACACA-----ACAGTACAGTACAAACAAGTACAGTACAGTGA 1175
Db 59 HsthrthtrthrglythrgasprlIethrleuserrhsgly----- 72
QY 1174 GAAAGACAAAACACACGAGCAAAACAAACACACACAGTGAATTCGAAAATATCAAC 1115
Db 73 -----HisthrlIethrhishetashtr----- 80
QY 1114 TGCGAATATGCAACCCCTTACATCCACACACAAATATGCAATATGCAATATGCAATG 1055
Db 81 -----ProthrhieCysenishetaserthrlatthrhishthralathre 95
QY 1054 GCATATATTCGAGCATGATAGTCTGCAATTCATTCACACATCTGCCAAACATATCTCAA 995
Db 96 Serhsglunishtr-----Serlle-Proserhishishis----- 107
QY 994 CACATGACCCCGATTCATGATACACACACATTCAT-----CGTTGCAATAG 944
Db 108 -----ThrhieCysenishvalasprthraerththtrhishar 118
QY 943 ATATTCATGAGTCTCTCTCCGACAGTGGTCAATGATCCATGCAATGCAATGCAATGCA 884
Db 118 g-----HisCysishthra 123
QY 883 TCCCATGATCTACATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 827
Db 123 sptnglnasnthValthrgarghnsishislaasprthProphronisgluhs 143
QY 826 ATGACCGAGATGGAATGACTGCTAGTACGACTGCAAGATGATCCAAAGATGAG 767
Db 143 ertharg-----Hisserlatatnglile-----Hisishisthrglum 157
QY 766 TTATATGACAGGCTCTGACCAATACACACACC 735
Db 157 etargthhisCysenishthraerththtr 167

RESULT 19
US-09-759-143-479
; Sequence 479, Application US/09759143
; Patent No. US20020022248A1
; GENERAL INFORMATION:
; APPLICANT: Xu, Jiangchun
; APPLICANT: Dillon, Davin C.
; APPLICANT: Mitcham, Jennifer L.
; APPLICANT: Harlocker, Susan L.
; APPLICANT: Jiang, Yugu
; APPLICANT: Henderson, Robert A.
; APPLICANT: Kalos, Michael D.
; APPLICANT: Retter, Marc W.
; APPLICANT: Fanger, Gary R.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Hepler, William T.
; APPLICANT: Day, Craig H.
; APPLICANT: Vedvick, Thomas S.
; APPLICANT: Carter, Derrick
; APPLICANT: Li, Samuel
; APPLICANT: Wang, Aijun
; APPLICANT: Fanger, Gary R.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
; TITLE OF INVENTION: DIAGNOSIS OF PROSTATE CANCER
; FILE REFERENCE: 210121.427C23
; CURRENT APPLICATION NUMBER: US/09/759, 143
; NUMBER OF SEQ ID NOS: 934
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 479
; LENGTH: 222
; TYPE: PRT
; ORGANISM: Homo sapiens

```

US-09-759-143-479

**Alignment Scores:**

Pred. No.:	0.0961	length:	2222
Score:	102.50	Matches:	60
Percent Similarity:	30.95%	Conservative:	18
Best Local Similarity:	23.81%	Mismatches:	63
Query Match:	3.84%	Indels:	111
DB:	10	Gaps:	14

US-10-020-441-1 (1-1477) x US-09-759-143-479 (1-222)

```

QY      1453 CACATCTTACGCACTGGCAGTAATCGAACAATGAATATGCGAGAAGTTTGCAAGTCGTAC   1394
           ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB      16 GlnSerHisgIuHsthrIgYLeval-----ThrTrp---ThrSp   28
QY      1393 ACCATGTGATGCAGC-----ACGAGTAATCGACCTTGAAGGATGGAGGAG   1346
           ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB      29 ThrGlnPrrTyrgYguUleThrlleThlHisHisIsthIlleThrgYlThrcLnlrthr   48
QY      1345 AGGTCCACCACTGTCACTGGAGCGGTAATCTGCCAAATTGTACTGCTTTTCACACA   1286
           ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB      49 HisglAspIleThrlThrlTrpLr-----                    56
QY      1285 TATAGATGGCATTTGGTGGCATGTGCGCATTCATTAATAGTGGTCCACCTGTGGNC   1226
DB      57 -----HisCys 58
QY      1225 AAACAAACAACACACAAA-----ACAGTACATAACAACAGTACGTACAGATGA   1175
           ||||| ||| ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB      59 HisthrTrlThrlngYlThrlAgAspIleThrlleuSenHisgl-----       72
QY      1174 GAAGAACAACAAACAACGGAACAAACAAACAAACAACAGTAGAATTTCGAAATATCAAC   1115
           ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB      73 -----HisThrIleThrlHisMetAsnThr-----            80
QY      1114 TGCGACAAATTCACCCCTTTTACATACTCCSACAACAATATGTCSATATCAATCATGT   1055
           ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB      81 -----ProThrHisCysHisMetAspThrAlaThrIsthIleThrlAlaThrlleu   95
QY      1054 GCACATATTCGACGAGTCAAGTCACTGTCCATGTCSACACACTGCCAACACATATCTCAA   995
           ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB      96 SerHisgYlHisThr-----Setile-ProSerHisShHis-----    107
QY      994 CACATTCGACCCCTCATTCATGTACACACACACTTCAT-----CGTTTGCATAG   944
           ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB      108 -----ThrHisCysHisValAspThrAlaThrIsthIleThrlAlaThrlleu   118
QY      943 ATATTTCATGAGTCTTCTCTCTCCGAAAGTGCTCATCATCCCATCAATACATTCACA   884
           ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB      118 G-----His-CysHisThrA-----His-CysHisThrA   123
QY      883 TCSCATCAATCTACCTCACTACCTCACATCAGCNC--ACCTCACTCAACAAACA   827
           ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB      123 spTHrGlAsnThValThrlArgRghIsHisHisAlAspThrPrOrPhHisgYlHis   143
QY      826 ATGACCGAGATGGAATCACTCTCTACTGAATGHTGCACAAATGTCAACAAATCAACAGAGTTG   767
           ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB      143 eThrlArg-----HiserLahrhrgInle-----HisHisIsthIngUm   157
QY      766 TTATTATCAAGGCTCTGACCAATCAACAACC   735
           ::::: ||::: ||||| |||||
DB      157 eLArgThrlHisCyshIsthIsthAspThrThrlThr   167

RESULT 20
; Sequence 479, Application US/09780669
; Patent No. US20020051977A1
; GENERAL INFORMATION:
; APPLICANT: Xu, Jiangchun
; APPLICANT: Dillon, Davin C.
; APPLICANT: Mitcham, Jennifer L.
; APPLICANT: Hartlocker, Susan L.
```

[illegible]

Db 108 -----ThrHisCysHisValAspThrArgThrHisArg 118  
Qy 943 ATATTTCATGAGTCTTCTCTCCGAGAGTGTCTCAATCATGCTCCATCATGATCAATGCTCA 884  
Db 118 g-----His-CysHisThr 123  
Qy 883 TCCCATCCAAATCTAATCTCAATGCTCAATGCTCAATGCTCAATGCTCAATGCTCAATGCTCA 827  
Db 123 sPTnGlnAsnThrValThrArgArgHisHisHisAlaAspThrProProHisGlyHis 143  
Qy 826 ATGACCGAGATGGAATCACTCTGCTACTGAACTGTGACAAAGTATCCAAACAGATTG 767  
Db 143 eTnThrArg-----HisSerAlaThrGlnIle-----HisHisThrGln 157  
Qy 766 TTATATCAAGGCTGTGACCAATACACACACC 735  
Db 157 eTArgThrHisCysHisThrAspThrThrThr 167

RESULT 21  
US-09-822-827-479  
; Sequence 479, Application US/09822827  
; Patent No. US20020081680A1  
; GENERAL INFORMATION:  
; APPLICANT: Xu, Jianshun  
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND  
; FILE REFERENCE: 210121.534C1  
; CURRENT APPLICATION NUMBER: US/09/822,827  
; NUMBER OF SEQ ID NOS: 982  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 479  
; LENGTH: 222  
; TYPE: PRP  
; ORGANISM: Homo sapiens  
US-09-822-827-479

Alignment Scores:  
Pred. No.: 0.0961 Length: 222  
Score: 102.50 Matches: 60  
Percent Similarity: 30.95% Conservative: 18  
Best Local Similarity: 23.81% Mismatches: 63  
Query Match: 3.84% Indels: 111  
DB: 10 Gaps: 14

US-10-020-441-1 (1-1477) x US-09-822-827-479 (1-222)

Qy 1453 CAACATTCCTAGCCACACTGGCATCTCGACAATGATATCGGGAAGTTTGGCAGTGTGAC 1394  
Db 16 GlnSerHisGlnHisThrGlyIleVal-----ThrTrp---ThrAsp 28  
Qy 1393 ACCATGTGATCGAC-----ACCGATACTGGAACCTTGAAGGATGGAGAG 1346  
Db 29 ThrGlnThrTyGlyGlnIleThrLeuThrHisHisThrIleThrGlyThrGlnThr 48  
Qy 1345 AGGTCAACCACTGTCACCTGGAGCGGTAAGTCGGAATTCGACCTGCTTACACA 1286  
Db 49 HisGlyAspIleThrThrTrpThr-----HisCys 58  
Qy 1285 TATAGATTCGCAATGTCGATGTCGCAATTCATTAATGCTGCTGACCTGTGTGC 1226  
Db 57 -----HisCys 58  
Qy 1225 AAACAAACACACACACA-----ACAGTACAGTACACACAGTACAGTACAGTAC 1175  
Db 59 HisThrThrThrGlyThrArgAspIleThrLeuSerHisGly-----HisCys 72  
Qy 1174 GAAGAAGACAAACACAGGACAAACAAACAAACAGTAAATTTGGAATATCAAC 1115  
Db 73 -----HisThrIleThrHisThrThrThr-----HisCys 80  
Qy 1114 TGGACAATATTCACCTTTTACATACCCACACAATAGTCCATATGATCAATGATCAATGT 1055

Db 81 -----ProThrHisCysHisValSerAspThrAlaThrHisThrLeu 95  
Qy 1054 GCACATATATGACAGATCAAGTCAATGCTCAATGCTCAATGCTCAATGCTCAATGCTCAATGCTCA 995  
Db 96 SerHisGlyHisThr-----SerIle-ProSerHisHisHis-----HisCys 107  
Qy 994 CACATGACCCCTCAATTCATGATGACAAACACACACTTCCAT-----CGTTGACATAG 944  
Db 108 -----ThrHisCysHisValAspThrArgThrHisArg 118  
Qy 943 ATATTTCATGAGTCTTCTCTCCGAGAGTGTCTCAATCATGCTCCATCATGATCAATGCTCA 884  
Db 118 g-----His-CysHisThr 123  
Qy 883 TCCCATCCAAATCTAATCTCAATGCTCAATGCTCAATGCTCAATGCTCAATGCTCAATGCTCA 827  
Db 123 sPTnGlnAsnThrValThrArgArgHisHisHisAlaAspThrProProHisGlyHis 143  
Qy 826 ATGACCGAGATGGAATCACTCTGCTACTGAACTGTGACAAAGTATCCAAACAGATTG 767  
Db 143 eTnThrArg-----HisSerAlaThrGlnIle-----HisHisThrGln 157  
Qy 766 TTATATCAAGGCTGTGACCAATACACACACC 735  
Db 157 eTArgThrHisCysHisThrAspThrThrThr 167

RESULT 22  
US-09-216-393-341  
; Sequence 341, Application US/09216393  
; Patent No. US20010014447A1  
; GENERAL INFORMATION:  
; APPLICANT: Milhausen, Michael James  
; TITLE OF INVENTION: TOXOPLASMA GONDII PROTEINS, NUCLEIC ACID MOLECULES, AND  
; FILE REFERENCE: TX-1-C2  
; CURRENT APPLICATION NUMBER: US/09/216,393  
; EARLIER FILING DATE: 1998-12-18  
; EARLIER APPLICATION NUMBER: 08/994,825  
; NUMBER OF SEQ ID NOS: 364  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 341  
; LENGTH: 288  
; TYPE: PRP  
; ORGANISM: Toxoplasma gondii  
US-09-216-393-341

Alignment Scores:  
Pred. No.: 0.105 Length: 288  
Score: 102.50 Matches: 51  
Percent Similarity: 39.38% Conservative: 25  
Best Local Similarity: 26.42% Mismatches: 82  
Query Match: 3.84% Indels: 35  
DB: 10 Gaps: 5

US-10-020-441-1 (1-1477) x US-09-216-393-341 (1-288)

Qy 1273 ATGGTGCATGTCGGCATTCATTTATATAGTCTGACACCTGTGTGCAAAACAACAC 1214  
Db 128 MetGlyGlnGlnAlaAlaPheIleArgProArgArgSerIleArgGlyValGlyThrThr 147  
Qy 1213 ACAACACAGTACAGTACAAAGTACAGTACAGTACAGTACAGTACAGTACAGTACAGTACAG 1154  
Db 148 ThrThrThr-SerSerSerThrThrThrThrThrThrThrThrThrThrThrThrThrThr 167  
Qy 1153 ACAAAACAAACACACAAAGTAAATTTGGAATATCAACTGACCAATATCCACCTTTA 1094  
Db 167 rThrThrThrThr-----ThrThrThrThrThrThrThrThrThrThrThrThrThrThr 181  
Qy 1093 CACTACCCACACAACATATGTCACATACATCAATGATGACATATGTCACATATGTCACATCAAG 1034  
Db 181 oThrThrThrThrThr-----ThrThrThr 189



```
; SOFTWARE: Annonmax Sequence Listing Engine vers. 1.1
; SEQ ID NO 37944
; LENGTH: 292
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
;   OTHER INFORMATION: MAP TO AC005414.2
;   OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 5.5
;   OTHER INFORMATION: EXPRESSED IN HBL100, SIGNAL = 6.2
;   OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 7.1
;   OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 6.1
;   OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 5.5
;   OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 8.2
;   OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 7.7
;   OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 3.2
;   OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 7.9
;   OTHER INFORMATION: EXPRESSED IN LONG, SIGNAL = 11
US-09-864-761-37944
```

```
Alignment Scores:
Pred. No.: 0.106 Length: 292
Score: 102.50 Matches: 51
Percent Similarity: 35.09% Conservative: 29
Best Local Similarity: 22.37% Mismatches: 86
Query Match: 3.84% Indels: 62
DB: 10 Gaps: 7
```

US-10-020-441-1 (1-1477) x US-09-864-761-37944 (1-292)

```
OY 1132 AATTTCGAATAATGACATGCGACAAATATCCACCCCTTTACCTACCCACAAACATAGT 1073
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 23 SerPheAlGhNIshNIshNIshNIshNIshNIshNIshNIshNIshNIshNIshNIsh 42
OY 1072 CCACATACATACATGCGACATATATTCGACGATCACTGCTCCATCTCCACCATC 1013
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 43 -HisTyrNIshNIshNIshNIshNIshNIshNIshNIshNIshNIshNIshNIshNI 62
OY 1012 TCCAAACACATACATGCAACATGACCCCTTCACTGATGTCACAAACATTCGCATCG 953
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 62 shNIshNIshNIshNIshNIshNIshNIshNIshNIshNIshNIshNIshNIshNIsh 76
OY 952 TTTCGATGATATTTTCATGATGATCTTCTCTCCCAAGTGGTCTCAATGCCATCAAT 893
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 76 shNIshNIshTyr-----HisNIshNIsh 83
OY 892 CACATTCATCCCATCCCATCTACTCACTCACTCACTCACTCACTCACTCACTCACTCA 833
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 83 sProhNIshNIsh-----His-AsnNIshNIsh-----HisNIshNIsh 94
OY 832 AACACATGACCGAGATGGAATACATCTGCTACTGATGACATGTCACAAAGTCATCCAA 773
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 94 ishNIshNIshNIshNIshTyrProthNIshNIshNIshNIshNIshNIshNIshNIshNI 113
OY 772 CAGTTGATTATGCAAGGCTGTGACCAATACAAACACCGGATGAGTACAGACAGACAGA 713
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 113 hrllethNIshNIshNIshNIshNIshNIshNIshNIshNIshNIshNIshNIshNIsh 133
OY 712 CACAAACGATACACAGACACAGACATGACTGATGATGATGATGATGATGATGATGATG 653
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 133 hrlNIshNIshNIshNIshNIshNIshNIshNIshNIshNIshNIshNIshNIshNIsh 152
OY 652 TGTGCTTATTC-----ACCAACACATCACTTCT 623
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 153 ThrNIshNIshNIshNIshNIshNIshNIshNIshNIshNIshNIshNIshNIshNIsh 172
OY 622 TCA----- 620
DB 173 ProPserGIshNIshNIshNIshNIshNIshNIshNIshNIshNIshNIshNIshNIsh 192
OY 619 ---AATATCCACATTTCTTACGATGACGATGACGATGATGATGATGATGATGATGATG 563
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 193 TyrLeuTyrTyrNIshNIshNIshNIshNIshNIshNIshNIshNIshNIshNIshNIsh 212
```

```
OY 562 CAATGAAAAACCCAGTTC 545
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 213 hIshNIshNIshNIshNIshNIshNIshNIshNIshNIshNIshNIshNIshNIsh 218
```

RESULT 25

US-09-801-368-274

; Sequence 274, Application US/09801368

; Patent No. US20020128250A1

; GENERAL INFORMATION:

; APPLICANT: Busby, Robert

; APPLICANT: Cali, Brian

; APPLICANT: Hecht, Peter

; APPLICANT: Holtzman, Doug

; APPLICANT: Madden, Kevin

; APPLICANT: Maxon, Mary

; APPLICANT: Milne, Todd

; APPLICANT: No. US20020128250A1man, Thea

; APPLICANT: Royer, John

; APPLICANT: Salama, Sofie

; APPLICANT: Sherman, Amir

; APPLICANT: Silva, Jeff

; APPLICANT: Summers, Eric

; TITLE OF INVENTION: Methods for Improving Secondary Metabolite Production in Fungi

; FILE REFERENCE: 109272.147

; CURRENT FILING DATE: 2001-03-07

; PRIOR FILING DATE: 2000-01-19

; PRIOR APPLICATION NUMBER: US 60/160,587

; PRIOR FILING DATE: 1999-10-20

; NUMBER OF SEQ ID NOS: 440

; SOFTWARE: PatentIn version 3.0

; SEQ ID NO 274

; LENGTH: 692

; TYPE: PRT

; ORGANISM: Saccharomyces cerevisiae

US-09-801-368-274

```
Alignment Scores:
Pred. No.: 0.178 Length: 692
Score: 101.50 Matches: 57
Percent Similarity: 37.50% Conservative: 51
Best Local Similarity: 19.79% Mismatches: 101
Query Match: 3.81% Indels: 79
DB: 10 Gaps: 8
```

US-10-020-441-1 (1-1477) x US-09-801-368-274 (1-692)

```
OY 1227 GCAACAAACACACACACACAGTACCAACGCTACAGTACAGTACAGTACAGTACAGT 1168
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 18 SerSerArgAnProArgProGlnValSerArgNIshNIshNIshNIshNIshNIshNIshNIsh 37
OY 1167 CAACACACACGAAACAAACAAACACACACAGTGAATTTGAAAATATCACTGGCACA 1108
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 38 AlAluYSerAnluYSerSerArgSerArgSerleuProSerSerSerThrNIshNIsh 57
OY 1107 AATCCACCCCTTTACATACCCACACACACATATGATGCCACATATCATCATGTCAGAT 1048
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 58 ThrAnSerAnValProAspProSerThrProSerleuProAnleuGlnValAnshNIsh 77
OY 1047 AATGACATAGTACGATCTCTCCATCTCACCCTTCCA-----AACACATATCA 997
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 78 GlnArgNIshSerSerNIshThrAnArgTyrNIshPheProSerSerSerNIshNIshNIsh 97
OY 996 AACACA-----TTGACCCCTTACATTCATGATGACACACACACACTTCATCGT 952
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 98 AsnSerGlnAnGlnLeuLeuThrThrProSerSerSerSerThlyArgProSerThr 117
OY 951 TTGCATATATTTCAATGATCTCTCCCTCC----- 919
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 118 SerArgArgSerSerTyrAnThrNIshNIshNIshNIshNIshNIshNIshNIshNIshNIsh 137
OY 918 -----AAGTGTCTCAATCATCCCATCATCATCATCATCATCATCATCATCATCAT 883
```

Db 138 MetGluProLysSerProIleuLysThrAsnAsnSerSerThrHisValSerLysHis 157  
Qy 882 CCCATTCATATCACTGACCTGACATGACGTCACCTGACACCAACAGACATGA 823  
Db 158 LysSerSerLysSerSerThrLysLysGluAsnAlaLeuThrAspAspAsnAspHis 177  
Qy 822 CCGAGATGGAATCACTGCTGCTACTGAA-----CTGTCAACAAGTCATCCAA 775  
Db 177 pLysAspAsnAspIleSerHisThrLysAspPheSerArgSerSerAsnSerArgProSe 197  
Qy 774 CACAGTT----- 768  
Db 197 rSerIleArgSerGlySerValSerArgLysSerAspValThrHisGluGluProHis 217  
Qy 767 -----GTTATATCAAGGCTGACCAATAC 742  
Db 217 nAsnGlySerLysSerSerAsnAsnGlnGluAsnTyrLeuValGlnAlaLeuThrArgSe 237  
Qy 741 AACAAAC-----GGATCAGTATGACAGAC 718  
Db 237 rAsnSerHisAlaSerSerLeuHisSerArgLysSerSerPheGlySerAspGlyAsnHis 257  
Qy 717 AGACACAAACAGATATACAGACACACAGATTAACCTGACATCTGACGAGACTG 658  
Db 257 rAlaTyrSerThrProLeuAsnSerProGlyLeuSer-Lys-----LeuThrAspH 274  
Qy 657 AGTCATGCTGTTATTCACCAACATGCTTCTTTCATATTCACCACTCT----- 605  
Db 274 iSerGlyGluTyrPheThrSerAsnSerThrSerLeuAsnHisHisSerSerArgA 294  
Qy 604 -----TACGTGACGATCAGC 590  
Db 294 sPLeTyrProSerLysHis 300

RESULT 26  
US-10-067-761-19  
; Sequence 19, Application US/10067761  
; Publication No. US20020197701A1  
; GENERAL INFORMATION:  
; APPLICANT: Shi et al.  
; TITLE OF INVENTION: Serine protease Polynucleotides, Polypeptides, and Antibodies  
; FILE REFERENCE: PTO05P4  
; CURRENT APPLICATION NUMBER: US/10/067,761  
; CURRENT FILING DATE: 2002-02-08  
; PRIOR APPLICATION NUMBER: 09/804,156  
; PRIOR FILING DATE: 2001-03-13  
; PRIOR APPLICATION NUMBER: 60/189,025  
; NUMBER OF SEQ ID NOS: 38  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 19  
; LENGTH: 238  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-10-067-761-19

Alignment Scores:  
Pred. No.: 0.156 Length: 238  
Score: 100.50 Matches: 49  
Percent Similarity: 37.71% Conservative: 17  
Best Local Similarity: 28.00% Mismatches: 50  
Query Match: 3.74% Indels: 59  
DB: 9 Gaps: 8

US-10-020-441-1 (1-1477) x US-10-067-761-19 (1-238)  
Qy 52 TTCGCGTTCATGCATTTCTTAACAGACAGAGACACA---ATGTGACAGSTTCACTAGTC 108  
Db 41 TTPPTGTGlnAlaSerLeuArgLeuArgValHisValCysGlyGlySerLeuLeu 60  
Qy 109 TCAACGAGACGACTCAACAGCTGATCATTTGTGTTCACCACTGSCAGATGATTCG 168

Db 61 SerProGlnTyrValLeuThrAlaAlaHisCysPheSerGlySerLeu-----Asn 77  
Qy 169 GTAAGAGATTCAGCTGAAACACATTTGTTGCTCAATGATTAATGATTTTCACAGACAGTG 228  
Db 78 SerSerAspTyrGlnValHisLeu----- 85  
Qy 229 TTTGCGTGTGCGTGGCTTTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 288  
Db 86 -----GlyGluLeuGluIleThrLeuSer 93  
Qy 289 ATATACCTTGCTTATTTTCAGGTTTCATTTCACACAGCAGGAATGCGACCAACAGGC 348  
Db 94 ProHis-----PheSerThrValArg-----GlnIleIle 103  
Qy 349 ATTCATCACCACCAACGCTGAGATTAAAGGTGACAGCAGATACATCCCTTTGTATGTGCG 408  
Db 104 LeuHisSerSerProSerGly-----GlnProGly----- 113  
Qy 409 GCACGACAGAGAGACCAATGCGACACAGACTGAGTGAATTCGATATTCGAAATTTGAATG 468  
Db 114 -----ThrSerGly-----AspIleAlaLeuValGlu 122  
Qy 469 CTGGCTCAATGCTGCACTTACAGAGTGAATCAGAGTGAATGAGTGCACAGGCATCG 528  
Db 123 LeuSerValProValThrLeuSerSerArgLysLeuProValCysLeuProGlnAlaSer 142  
Qy 529 GATATCCCGCCACCTGCACTGCTGTTTCATTTGTTGTTGTTGTTGTTGTTGTTGTTGTT 573  
Db 143 AspAspPheCysProGlyIleArgCysTyrPvalThrGlyTyrPglY 157

RESULT 27  
US-09-804-156-19  
; Sequence 19, Application US/09804156  
; Patent No. US20020068320A1  
; GENERAL INFORMATION:  
; APPLICANT: Shi et al.  
; TITLE OF INVENTION: Serine protease Polynucleotides, Polypeptides, and Antibodies  
; FILE REFERENCE: PTO05P4  
; CURRENT APPLICATION NUMBER: US/09/804,156  
; CURRENT FILING DATE: 2001-03-13  
; PRIOR APPLICATION NUMBER: 60/189,025  
; PRIOR FILING DATE: 2000-03-14  
; NUMBER OF SEQ ID NOS: 38  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 19  
; LENGTH: 238  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-804-156-19

Alignment Scores:  
Pred. No.: 0.156 Length: 238  
Score: 100.50 Matches: 49  
Percent Similarity: 37.71% Conservative: 17  
Best Local Similarity: 28.00% Mismatches: 50  
Query Match: 3.74% Indels: 59  
DB: 10 Gaps: 8

US-10-020-441-1 (1-1477) x US-09-804-156-19 (1-238)  
Qy 52 TTCGCGTTCATGCATTTCTTAACAGACAGAGACACA---ATGTGACAGSTTCACTAGTC 108  
Db 41 TTPPTGTGlnAlaSerLeuArgLeuArgValHisValCysGlyGlySerLeuLeu 60  
Qy 109 TCAACGAGACGCTATCAACAGCTGATCATTTGTGTTCACCACTGSCAGATGATTCG 168  
Db 61 SerProGlnTyrValLeuThrAlaAlaHisCysPheSerGlySerLeu-----Asn 77  
Qy 169 GTAAGAGATTCAGCTGAAACACATTTGTTGCTCAATGATTAATGATTTTCACAGACAGTG 228  
Db 78 SerSerAspTyrGlnValHisLeu----- 85  
Qy 229 TTTGCGTGTGCGTGGCTTTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 288

Db 86 -----GlyGluLeuGlnLeuThrLeuSer 93  
Qy 289 ATATCTGTCTTATTTCAGGTTTCATTTCTACACCTGAGGAATGGACACAAAGGC 348  
Db 94 ProHis-----PheSerThrValArg-----GlnIleIle 103  
Qy 349 ATCCATCAACCAACCGTCTGAGTTAAGTGGACACAGATACATGCGCTTGTATATGCG 408  
Db 104 LeuHisSerSerProSerIly-----GlnProGly----- 113  
Qy 409 GCACACACAGAGAGACCAATCCGACAGACATCAGTGGATTGGATTGCAATTGTAATG 468  
Db 114 -----ThiSerGly-----AspIleAlaLeuValGln 122  
Qy 469 CTGGCTCAATGTGTAACCTTACAGAGTGAATCAGATGATCAGTCTGCACAGCCATG 528  
Db 123 LeuSerValProValThrLeuSerSerArgIleLeuProValCysLeuProGluAlaSer 142  
Qy 529 GATATCCGCGCACCTGGAACCTGATGTTTTCATTTGTTGTTATGGA 573  
Db 143 AspArgPheCysProGlyIleArgCysTrpValThrGlyTrpGly 157  
RESULT 28  
US-10-117-323-37  
; Sequence 37, Application US/10117323  
; Publication No. US2003005493A1  
; GENERAL INFORMATION:  
; APPLICANT: Rancourt, Derrick E.  
; APPLICANT: Rancourt, Susan L.  
; APPLICANT: O'Sullivan, Colleen M.  
; TITLE OF INVENTION: Implantation Serine Proteases  
; FILE REFERENCE: 033337-005  
; CURRENT APPLICATION NUMBER: US/10/117,323  
; CURRENT FILING DATE: 2002-04-08  
; PRIOR APPLICATION NUMBER: US 60/281,724  
; PRIOR FILING DATE: 2001-04-06  
; PRIOR APPLICATION NUMBER: US 60/294,736  
; PRIOR FILING DATE: 2001-05-30  
; PRIOR APPLICATION NUMBER: US 60/350,962  
; PRIOR FILING DATE: 2002-01-25  
; NUMBER OF SEQ ID NOS: 41  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 37  
; LENGTH: 272  
; TYPE: PRT  
; ORGANISM: Mouse  
US-10-117-323-37  
Alignment Scores:  
Pred. No.: 0.23 Length: 272  
Score: 99.00 Matches: 22  
Percent Similarity: 59.26% Conservative: 10  
Best Local Similarity: 40.74% Mismatches: 20  
Query Match: 3.69% Indels: 2  
DB: 9 Gaps: 1  
US-10-020-441-1 (1-1477) x US-10-117-323-37 (1-272)  
Qy 427 ATCGACAGACACTGAGTTCGATTCATATTCGATTCGCTGCTCAATGTGTAAC 486  
Db 112 IleAlaGln-----AspGlyAlaAspIleAlaLeuLeuLeuThrAsnProValAsn 129  
Qy 487 TTACGAGTGGAAATCAGAGTATCAGTCTGCCACAGCCATCGGATATCCCGCACCTGGA 546  
Db 130 IleThrSerAsnValHisThrValSerLeuProProAlaSerGlnThrPheProSerGly 149  
Qy 547 ACTGCTGTTTCATGTTGTTATGGAAGAGATGAATACGAC 588  
Db 150 ThrLeuCysTrpValThrGlyTrpGlyAsnIleAsnAsp 163  
RESULT 29  
US-10-041-006A-7

; Sequence 7, Application US/10041006A  
; Patent No. US20020168754A1  
; GENERAL INFORMATION:  
; APPLICANT: Andrew-Gordon, Patricia  
; APPLICANT: Darrow, Andrew  
; APPLICANT: Qi, Jian-shen  
; TITLE OF INVENTION: DNA encoding the novel human serine  
; FILE REFERENCE: ORT-1032  
; CURRENT APPLICATION NUMBER: US/10/041,006A  
; CURRENT FILING DATE: 2002-01-07  
; NUMBER OF SEQ ID NOS: 11  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 7  
; LENGTH: 290  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-10-041-006A-7  
Alignment Scores:  
Pred. No.: 0.235 Length: 290  
Score: 99.00 Matches: 49  
Percent Similarity: 36.90% Conservative: 20  
Best Local Similarity: 26.20% Mismatches: 48  
Query Match: 3.69% Indels: 70  
DB: 9 Gaps: 9  
US-10-020-441-1 (1-1477) x US-10-041-006A-7 (1-290)  
Qy 91 TGTACAGSTTCACTAGTCTCAACAGAGACAGTACTCAGCTGCTGATCTGTTGTTCTCA 150  
Db 60 CysGlyGlySerLeuIleAlaGlnIntrValIleThrAlaIleHisCysPhe----- 77  
Qy 151 CCATTGCCAGTATTCGGGTGAAGATGAGATGAGTGAACATTTGTTGTCATATGTAATGA 210  
Db 78 -----Arg 78  
Qy 211 TTGATTTACAGACAGTGTTCGTCGTCGTCGTTGTTGTTGTAAGCAGTGGTTT 270  
Db 79 AsnThrSerGluThr----- 83  
Qy 271 GCATGATTTTAATGCTATATCTGATCTGTTGTTATTTACAGTTTCATTTCTACACTGAG 330  
Db 84 -----SerLeuTrpGlnValLeuLeu----- 90  
Qy 331 AATGCGACCAACAGGACATCCATCACCACCGCTGAGTTAAGTGGACACAGATAC 390  
Db 91 ---GlyAlaArgGln-----LeuValGlnProGly--- 99  
Qy 391 ATGCCCTCTTGTATGTCGACACAGACAGAGAGA-----CCAATGCAACAG--- 435  
Db 100 ---ProHisAlaMetTrpAlaArgValArgGlnValGluSerAsnProLeuTrpGlnGly 118  
Qy 436 ACACCTCAGTGTGATTCGATTCGATTCGATTCGATTCGATTCGATTCGATTCGATTCGAT 495  
Db 119 ThrAlaSerSerAlaAspValAlaLeuValGluLeuGlnAlaProValProPheThrAsn 138  
Qy 496 GGAATCAAGATGATCAGTCTGCCACAGCCATCGATATCCCGCACCTGGAATGCTGTT 555  
Db 139 TyrIleLeuProValCysLeuProAspProSerValIlePheGluThrGlyMetAsnCys 158  
Qy 556 TTCATTTGTTGTTATGGAAGS---GATGATAACGACCGTGAATCCGTCACGTAAGAAATGT 612  
Db 159 TrpValThrGlyTrpGlySerProSerGlnGluAspLeuLeuProGluProArg----- 176  
Qy 613 GGAATATTGAAGAAATGAGT 633  
Db 177 ---IleLeuGlnIlyLeuAla 182  
RESULT 30  
US-10-028-072-222  
; Sequence 222, Application US/10028072  
; Publication No. US2003000431A1

GENERAL INFORMATION:  
APPLICANT: Baker, Kevin P.  
APPLICANT: Beresini, Maureen  
APPLICANT: DeForge, Laura  
APPLICANT: Desnoyers, Luc  
APPLICANT: Filvaroff, Ellen  
APPLICANT: Geo, Wei-Qiang  
APPLICANT: Gerritsen, Mary E.  
APPLICANT: Goddard, Audrey  
APPLICANT: Godowski, Paul J.  
APPLICANT: Gurney, Austin L.  
APPLICANT: Sherwood, Steven  
APPLICANT: Smith, Victoria  
APPLICANT: Stewart, Timothy A.  
APPLICANT: Tumas, Daniel  
APPLICANT: Watanabe, Colin K  
APPLICANT: Wood, William  
APPLICANT: Zhang  
TITLE OF INVENTION:  
FILE REFERENCE:  
CURRENT APPLICATION NUMBER: US/10/028,072  
CURRENT FILING DATE: 2001-12-19  
PRIOR APPLICATION NUMBER: 60/049911  
PRIOR FILING DATE: 1997-06-18  
PRIOR APPLICATION NUMBER: 60/056974  
PRIOR FILING DATE: 1997-08-26  
PRIOR APPLICATION NUMBER: 60/059113  
PRIOR FILING DATE: 1997-09-17  
PRIOR APPLICATION NUMBER: 60/059115  
PRIOR FILING DATE: 1997-09-17  
PRIOR APPLICATION NUMBER: 60/059117  
PRIOR FILING DATE: 1997-09-17  
PRIOR APPLICATION NUMBER: 60/059122  
PRIOR FILING DATE: 1997-09-17  
PRIOR APPLICATION NUMBER: 60/059184  
PRIOR FILING DATE: 1997-09-17  
PRIOR APPLICATION NUMBER: 60/059263  
PRIOR FILING DATE: 1997-09-18  
PRIOR APPLICATION NUMBER: 60/059352  
PRIOR FILING DATE: 1997-09-19  
PRIOR APPLICATION NUMBER: 60/059588  
PRIOR FILING DATE: 1997-09-19  
PRIOR APPLICATION NUMBER: 60/059836  
PRIOR FILING DATE: 1997-09-24  
PRIOR APPLICATION NUMBER: 60/062250  
PRIOR FILING DATE: 1997-10-17  
PRIOR APPLICATION NUMBER: 60/062285  
PRIOR FILING DATE: 1997-10-17  
PRIOR APPLICATION NUMBER: 60/062287  
PRIOR FILING DATE: 1997-10-17  
PRIOR APPLICATION NUMBER: 60/062814  
PRIOR FILING DATE: 1997-10-24  
PRIOR APPLICATION NUMBER: 60/062816  
PRIOR FILING DATE: 1997-10-24  
PRIOR APPLICATION NUMBER: 60/063045  
PRIOR FILING DATE: 1997-10-24  
PRIOR APPLICATION NUMBER: 60/063082  
PRIOR FILING DATE: 1997-10-31  
PRIOR APPLICATION NUMBER: 60/063127  
PRIOR FILING DATE: 1997-10-24  
PRIOR APPLICATION NUMBER: 60/063327  
PRIOR FILING DATE: 1997-10-27  
PRIOR APPLICATION NUMBER: 60/063329  
PRIOR FILING DATE: 1997-10-27  
PRIOR APPLICATION NUMBER: 60/063550  
PRIOR FILING DATE: 1997-10-28  
PRIOR APPLICATION NUMBER: 60/063561  
PRIOR FILING DATE: 1997-10-28  
PRIOR APPLICATION NUMBER: 60/063704  
PRIOR FILING DATE: 1997-10-29  
PRIOR APPLICATION NUMBER: 60/063733  
PRIOR FILING DATE: 1997-10-29  
PRIOR APPLICATION NUMBER: 60/063735

PRIOR FILING DATE: 1997-10-29  
PRIOR APPLICATION NUMBER: 60/063738  
PRIOR FILING DATE: 1997-10-29  
PRIOR APPLICATION NUMBER: 60/063755  
PRIOR FILING DATE: 1997-10-17  
PRIOR APPLICATION NUMBER: 60/064248  
PRIOR FILING DATE: 1997-11-03  
PRIOR APPLICATION NUMBER: 60/064809  
PRIOR FILING DATE: 1997-11-07  
PRIOR APPLICATION NUMBER: 60/065186  
PRIOR FILING DATE: 1997-11-12  
PRIOR APPLICATION NUMBER: 60/065846  
PRIOR FILING DATE: 1997-11-17  
PRIOR APPLICATION NUMBER: 60/066364  
PRIOR FILING DATE: 1997-11-21  
PRIOR APPLICATION NUMBER: 60/066453  
PRIOR FILING DATE: 1997-11-24  
PRIOR APPLICATION NUMBER: 60/066511  
PRIOR FILING DATE: 1997-11-24  
PRIOR APPLICATION NUMBER: 60/066770  
PRIOR FILING DATE: 1997-11-24  
PRIOR APPLICATION NUMBER: 60/069212  
PRIOR FILING DATE: 1997-12-11  
PRIOR APPLICATION NUMBER: 60/069278  
PRIOR FILING DATE: 1997-12-11  
PRIOR APPLICATION NUMBER: 60/069334  
PRIOR FILING DATE: 1997-12-11  
PRIOR APPLICATION NUMBER: 60/069694  
PRIOR FILING DATE: 1997-12-16  
PRIOR APPLICATION NUMBER: 60/072320  
PRIOR FILING DATE: 1998-01-23  
PRIOR APPLICATION NUMBER: 60/073612  
PRIOR FILING DATE: 1998-02-04  
PRIOR APPLICATION NUMBER: 60/074086  
PRIOR FILING DATE: 1998-02-09  
PRIOR APPLICATION NUMBER: 60/074092  
PRIOR FILING DATE: 1998-02-09  
PRIOR APPLICATION NUMBER: 60/077791  
PRIOR FILING DATE: 1998-03-12  
PRIOR APPLICATION NUMBER: 60/078910  
PRIOR FILING DATE: 1998-03-20  
PRIOR APPLICATION NUMBER: 60/079294  
PRIOR FILING DATE: 1998-03-25  
PRIOR APPLICATION NUMBER: 60/079663  
PRIOR FILING DATE: 1998-02-27  
PRIOR APPLICATION NUMBER: 60/079728  
PRIOR FILING DATE: 1998-03-27  
PRIOR APPLICATION NUMBER: 60/080165  
PRIOR FILING DATE: 1998-03-31  
PRIOR APPLICATION NUMBER: 60/081203  
PRIOR FILING DATE: 1998-04-09  
PRIOR APPLICATION NUMBER: 60/081229  
PRIOR FILING DATE: 1998-04-09  
PRIOR APPLICATION NUMBER: 60/081695  
PRIOR FILING DATE: 1998-04-14  
PRIOR APPLICATION NUMBER: 60/081817  
PRIOR FILING DATE: 1998-04-15  
PRIOR APPLICATION NUMBER: 60/081818  
PRIOR FILING DATE: 1998-04-15  
PRIOR APPLICATION NUMBER: 60/082999  
PRIOR FILING DATE: 1998-04-24  
PRIOR APPLICATION NUMBER: 60/083322  
PRIOR FILING DATE: 1998-04-28  
PRIOR APPLICATION NUMBER: 60/083545  
PRIOR FILING DATE: 1998-04-29  
PRIOR APPLICATION NUMBER: 60/084600  
PRIOR FILING DATE: 1998-05-07  
PRIOR APPLICATION NUMBER: 60/084627  
PRIOR FILING DATE: 1998-05-07  
PRIOR APPLICATION NUMBER: 60/084637  
PRIOR FILING DATE: 1998-05-07  
PRIOR APPLICATION NUMBER: 60/085149  
PRIOR FILING DATE: 1998-05-12



```
/ PRIOR APPLICATION NUMBER: 60/085323
/ PRIOR FILING DATE: 1998-05-13
/ PRIOR APPLICATION NUMBER: 60/085338
/ PRIOR FILING DATE: 1998-05-13
/ PRIOR APPLICATION NUMBER: 60/085339
/ PRIOR FILING DATE: 1998-05-13
/ PRIOR APPLICATION NUMBER: 60/085579
/ PRIOR FILING DATE: 1998-05-15
/ PRIOR APPLICATION NUMBER: 60/085697
/ PRIOR FILING DATE: 1998-05-15
/ PRIOR APPLICATION NUMBER: 60/085704
/ PRIOR FILING DATE: 1998-05-15
/ PRIOR APPLICATION NUMBER: 60/086114
/ PRIOR FILING DATE: 1998-05-22
/ PRIOR APPLICATION NUMBER: 60/086430
/ PRIOR FILING DATE: 1998-05-22
/ PRIOR APPLICATION NUMBER: 60/087106
/ PRIOR FILING DATE: 1998-05-28
/ PRIOR APPLICATION NUMBER: 60/088026
/ PRIOR FILING DATE: 1998-06-04
/ PRIOR APPLICATION NUMBER: 60/088730
/ PRIOR FILING DATE: 1998-06-10
/ PRIOR APPLICATION NUMBER: 60/088741
/ PRIOR FILING DATE: 1998-06-10
/ PRIOR APPLICATION NUMBER: 60/088810
/ PRIOR FILING DATE: 1998-06-10
/ PRIOR APPLICATION NUMBER: 60/088858
/ PRIOR FILING DATE: 1998-06-11
/ PRIOR APPLICATION NUMBER: 60/089532
/ PRIOR FILING DATE: 1998-06-17
/ PRIOR APPLICATION NUMBER: 60/089599
/ PRIOR FILING DATE: 1998-06-17
/ PRIOR APPLICATION NUMBER: 60/089907
/ PRIOR FILING DATE: 1998-06-18
/ PRIOR APPLICATION NUMBER: 60/089947
/ PRIOR FILING DATE: 1998-06-19
/ PRIOR APPLICATION NUMBER: 60/090349
/ PRIOR FILING DATE: 1998-06-23
/ PRIOR APPLICATION NUMBER: 60/090429
/ PRIOR FILING DATE: 1998-06-24
/ PRIOR APPLICATION NUMBER: 60/090445
/ PRIOR FILING DATE: 1998-06-24
/ PRIOR APPLICATION NUMBER: 60/090538
/ PRIOR FILING DATE: 1998-06-24
/ PRIOR APPLICATION NUMBER: 60/090863
/ PRIOR FILING DATE: 1998-06-26
/ PRIOR APPLICATION NUMBER: 60/091360
/ PRIOR FILING DATE: 1998-07-01
/ PRIOR APPLICATION NUMBER: 60/091519
/ PRIOR FILING DATE: 1998-07-02
/ PRIOR APPLICATION NUMBER: 60/091982
/ PRIOR FILING DATE: 1998-07-07
```

## Alignment Scores:

```
Pred. No.: 0.235 Length: 290
Score: 99.00 Matches: 49
Percent Similarity: 36.90% Conservative: 20
Best Local Similarity: 26.20% Mismatches: 48
Query Match: 3.69% Indels: 70
DB: 9 Gaps: 9
```

US-10-020-441-1 (1-1477) x US-10-028-072-222 (1-290)

```
QY 91 TGTACAGTCTCACTGCTCAACGAGAGACTGACGCTGCTGCTGCTGCTCA 150
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 60 CysglyglyserleuileagluginleuValleuthralalahlscysphe----- 77
QY 151 CCATGCCAGTGAATTCGGGTAGAGAGTCACTGAACACATTGTTGCTCAATGTAATGCA 210
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 78 -----Atg 78
QY 211 TTGATTTCACAGACAGTGTTCGCTGTCGCTGCTGTTGTTGTGTATGACAGTGGCTTT 270
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
```

```
Db 79 AanthrSergluThr----- 83
QY 271 GCATGATTTTAATGCTATTAATCTTGCTTATTTACAGTTTCATTTCTACACTGAGG 330
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 84 -----SerleuTrglnValleu----- 90
QY 331 AATGGCAGCAACCAAGGATCCATCCACCAACCGTGTGAGTTAAGTGGCACCAGATAC 390
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 91 ---Glyalargln-----LeuValGlnProGly---- 99
QY 391 ATGCCCTCTTGATGTCCGACAGACAGAGAGA-----CCATGCCACAG----- 435
Db 100 ---ProHsalameLtyrAlarGvalArgValArgGlnValGuseAsnProLeuTrglnGly 118
QY 436 ACACTCAGTGATTCGATTTGCAATTTGAATGCTGCTCAATAGTCACTTACAGAGT 495
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 119 ThrAlaSerSerAlaAspValAlaLeuValGlnLeuGlnAlaProValProPheThrAsn 138
QY 436 GGAATCAGAGTGAATCAGTCCACAGCCATCGATATCCGCCACCTGGAATGTGTGT 555
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 139 TyrIleLeuProValCysLeuProAspProSerValIlePheGluThrGlyMetAsnCys 158
QY 556 TTCATTTGTTGCTTATGAGAG---GATGATTAACGACCGGTATCCGTCACTRAAGATGCT 612
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 159 TrpValTrnGlyTrpGlySerProSerGluGlnAspLeuProGluProArg----- 176
QY 613 GGAATTTGAAGAAAGTGAGT 633
Db 177 ---IleLeuGlnLysLeuAla 182
```

## RESULT 31

US-10-121-049-222

Sequence 222, Application US//10121049

Publication No. US20030022239A1

GENERAL INFORMATION:

```
/ APPLICANT: Baker, Kevin P.
/ APPLICANT: Beresini, Maureen
/ APPLICANT: Deforge, Laura
/ APPLICANT: Desnoyers, Luc
/ APPLICANT: Filvaroff, Ellen
/ APPLICANT: Gao, Wei-Qiang
/ APPLICANT: Gerritsen, Mary E.
/ APPLICANT: Goddard, Audrey
/ APPLICANT: Godowski, Paul J.
/ APPLICANT: Gurney, Austin L.
/ APPLICANT: Sherwood, Steven
/ APPLICANT: Smith, Victoria
/ APPLICANT: Stewart, Timothy A.
/ APPLICANT: Tumas, Daniel
/ APPLICANT: Watanabe, Colin K
/ APPLICANT: Wood, William
/ APPLICANT: Zhang, Zemin
/ TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
/ FILE REFERENCE: P3330R1C17
/ CURRENT APPLICATION NUMBER: US//10/121,049
/ CURRENT FILING DATE: 2002-04-12
/ Prior Application removed - See File Wrapper or Palm
/ NUMBER OF SEQ ID NOS: 550
/ SEQ ID NO 222
/ LENGTH: 290
/ TYPE: PRT
/ ORGANISM: Homo Sapien
US-10-121-049-222
```

## Alignment Scores:

```
Pred. No.: 0.235 Length: 290
Score: 99.00 Matches: 49
Percent Similarity: 36.90% Conservative: 20
Best Local Similarity: 26.20% Mismatches: 48
Query Match: 3.69% Indels: 70
DB: 9 Gaps: 9
```

```
US-10-020-441-1 (1-1477) x US-10-121-049-222 (1-290)
QY 91 TGTACAGSTTCACTAGTCTCAACAGAGACGACTACAGCTGTCATTGTTGCTCA 150
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 60 CysGlyGlySerLeuIleAlaGluGlnTrpValLeuThrAlaAlaHisCysPhe----- 77
QY 151 CCATTGCCAGTGAATTCGGGTAAAGAGATGCACTGAAACACATTGTTGCTCAATGTAATGCA 210
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 78 -----Arg 78
QY 211 TTGATTTCACAGACAGTGTTCGCGTGCCTGCGTTTGTTGTTGATGACAGTGGGTTT 270
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 79 AsnThrSerGluThr----- 83
QY 271 GCATGATTTTAATGCTATATCTTATCTGCTTATTTCAGTTTCATTTCACACTGAGG 330
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 84 -----SerLeuTrpGlnValLeuLeu----- 90
QY 331 AATGCCGACCAACAAAGGCATTCATCACCACCGTCTGAGTTAAGTGGCACCCAGAGATAC 390
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 91 --GlyAlaArgGln-----LeuValGlnProGly--- 99
QY 391 ATGCCCTCTTGATGTCCGACAGACAGAGAGA-----CCAATGCCACAG--- 435
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 100 --ProHsAlaMetTrpAlaArgValAlaArgGlnValGluSerAsnProLeuTrpGlnGly 118
QY 436 ACACCTCAGTGATTCGATATTCGCAATTGTAATGCTGCTCAATAGTCAACTTACAGAGT 495
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 119 ThrAlaSerSerAlaAspValAlaLeuValGluLeuGlnAlaProValProPheThrAsn 138
QY 496 GGAATCAGAGTATCATGTCCTCCACAGCCATGATATCCCGCCACCTGGAACTGGTGT 555
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 139 TyrIleLeuProValCysLeuProAspProSerValIlePheGluThrGlyMetAsnCys 158
QY 556 TTCATTGTTGGTTATGAGAAG---GATGATAACGACCGGATCCGACAGTAAAGATGT 612
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 159 TrpValThrGlyTrpGlySerProSerGluGluAspLeuLeuProGluProArg----- 176
QY 613 GGAATATTGAAGAAAGTAGT 633
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 177 ---IleLeuGlnIlybLeuAla 182

RESULT 32
US-10-123-904-222
; Sequence 222, Application US/10123904
; Publication No. US20030022328A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Beresini, Maureen
; APPLICANT: DeForge, Laura
; APPLICANT: Desnoyers, Luc
; APPLICANT: Filvaroff, Ellen
; APPLICANT: Gao, Wei-Qiang
; APPLICANT: Gerritsen, Mary E.
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Sherwood, Steven
; APPLICANT: Smith, Victoria
; APPLICANT: Stewart, Timothy A.
; APPLICANT: Tumas, Daniel
; APPLICANT: Wood, William
; APPLICANT: Wood, William K
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; FILE REFERENCE: P3330R1C54
; CURRENT FILING DATE: 2002-04-16
; Prior Application removed - See File Wrapper or Palm
; NUMBER OF SEQ ID NOS: 550
; SEQ ID NO 222
; LENGTH: 290
```

```
; TYPE: PRT
; ORGANISM: Homo Sapien
US-10-123-904-222
Alignment Scores:
Pred. No.: 0.235
Score: 99.00
Percent Similarity: 36.90%
Best Local Similarity: 26.20%
Query Match: 3.69%
DB: 9 Gaps: 9

US-10-020-441-1 (1-1477) x US-10-123-904-222 (1-290)
QY 91 TGTACAGSTTCACTAGTCTCAACAGAGACGACTACAGCTGTCATTGTTGCTCA 150
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 60 CysGlyGlySerLeuIleAlaGluGlnTrpValLeuThrAlaAlaHisCysPhe----- 77
QY 151 CCATTGCCAGTGAATTCGGGTAAAGAGATGCACTGAAACACATTGTTGCTCAATGTAATGCA 210
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 78 -----Arg 78
QY 211 TTGATTTCACAGACAGTGTTCGCGTGCCTGCGTTTGTTGTTGATGACAGTGGGTTT 270
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 79 AsnThrSerGluThr----- 83
QY 271 GCATGATTTTAATGCTATATCTTATCTGCTTATTTCAGTTTCATTTCACACTGAGG 330
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 84 -----SerLeuTrpGlnValLeuLeu----- 90
QY 331 AATGCCGACCAACAAAGGCATTCATCACCACCGTCTGAGTTAAGTGGCACCCAGAGATAC 390
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 91 --GlyAlaArgGln-----LeuValGlnProGly--- 99
QY 391 ATGCCCTCTTGATGTCCGACAGACAGAGAGA-----CCAATGCCACAG--- 435
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 100 --ProHsAlaMetTrpAlaArgValAlaArgGlnValGluSerAsnProLeuTrpGlnGly 118
QY 436 ACACCTCAGTGATTCGATATTCGCAATTGTAATGCTGCTCAATAGTCAACTTACAGAGT 495
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 119 ThrAlaSerSerAlaAspValAlaLeuValGluLeuGlnAlaProValProPheThrAsn 138
QY 496 GGAATCAGAGTATCATGTCCTCCACAGCCATGATATCCCGCCACCTGGAACTGGTGT 555
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 139 TyrIleLeuProValCysLeuProAspProSerValIlePheGluThrGlyMetAsnCys 158
QY 556 TTCATTGTTGGTTATGAGAAG---GATGATAACGACCGGATCCGACAGTAAAGATGT 612
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 159 TrpValThrGlyTrpGlySerProSerGluGluAspLeuLeuProGluProArg----- 176
QY 613 GGAATATTGAAGAAAGTAGT 633
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 177 ---IleLeuGlnIlybLeuAla 182

RESULT 33
US-10-140-470-222
; Sequence 222, Application US/10140470
; Publication No. US20030022331A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Beresini, Maureen
; APPLICANT: DeForge, Laura
; APPLICANT: Desnoyers, Luc
; APPLICANT: Filvaroff, Ellen
; APPLICANT: Gao, Wei-Qiang
; APPLICANT: Gerritsen, Mary E.
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Sherwood, Steven
; APPLICANT: Smith, Victoria
; APPLICANT: Stewart, Timothy A.
; APPLICANT: Tumas, Daniel
```

```

; APPLICANT: Watanabe, Colin K
; APPLICANT: Wood, William
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; FILE REFERENCE: P3330R1C160
; CURRENT APPLICATION NUMBER: US/10/140,470
; CURRENT FILING DATE: 2002-05-06
; Prior Application removed - See Palm or File Wrapper
; NUMBER OF SEQ ID NOS: 550
; SEQ ID NO 222
; LENGTH: 290
; TYPE: PRT
; ORGANISM: Homo Sapien
US-10-140-470-222

Alignment Scores:
Pred. No.: 0.235 Length: 290
Score: 99.00 Matches: 49
Percent Similarity: 36.90% Conservative: 20
Best Local Similarity: 26.20% Mismatches: 48
Query Match: 3.69% Indels: 70
DB: Gaps: 9

US-10-020-441-1 (1-1477) x US-10-140-470-222 (1-290)
OY 91 TGTACAGSTTTCAGTCTGCTCAACGAGACGATCTCAGCTGCTGCTATTTGCTCTCA 150
Db 60 CysGlyGlySerLeuIleAlaGluGlnTrpValLeuThrAlaAlaHisCysPhe----- 77
OY 151 CCATTGCCAGTGATTCGGGTGAAGATGACATGCAAAACATTTGTTCTCATGTATCGA 210
Db 78 ----- 78
OY 211 TTGATTTACAGACAGCTGTTTCCGTGCGTGGCGTTTGTGTTGTATGACGTGGGTTT 270
Db 79 AsnThrSerGluThr----- 83
OY 271 GCATGATTTTAATGCTATATATCTTGGTCTTATTTCAAGTTTCAATTTCTCAGCTGAGG 330
Db 84 -----SerLeuTrpGlnValLeu----- 90
OY 331 AATGGCGAACCAAGGATCCATCACCACCGTCTGAGTTAAGGTGCACACGAGATAC 390
Db 91 ---GlyAlaArgGln-----LeuValGlnProGly--- 99
OY 391 ATGCCCTCTTGTATGTCGGCAGCAGAGAGA-----CCATTCGCACAG--- 435
Db 100 ---ProHisAlaMetTyrAlaArgValArgGlnValGluSerAsnProLeuTyrGlnGly 118
OY 436 ACACTCAGTGGATTTGATATGCAATTGTAATGCTGGCTCAATAGTGTCACTTACAGAGT 495
Db 119 ThrAlaSerSerAlaAspValAlaLeuValGluLeuGlnAlaProValProPheThrAsn 138
OY 496 GGAATCAGAGTATGATGCTGCGCACAGCATCGATATCCGCCACCTGAGACTGGTGT 555
Db 139 TyrIleLeuProValCysLeuProAspProSerValIlePheGluThrGlyMetCys 158
OY 556 TTCATTGTTGGTTATGAGAG---GATGATTAACGACCGTATCCGTCAACGTAAAGATGCT 612
Db 159 TrpValThrGlyTyrGlySerProSerGluGluAspLeuLeuProGluProArg----- 176
OY 613 GGAATATTGAGAGAAAGTAGT 633
Db 177 ---IleLeuGlnLysLeuAla 182

RESULT 34
US-10-175-746-222
; Sequence 222, Application US/10175746
; Publication No. US20030027270A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Beresini, Maureen
```

```

; APPLICANT: DeForge, Laura
; APPLICANT: Desnoyers, Luc
; APPLICANT: Filvaroff, Ellen
; APPLICANT: Gao, Wei-Qiang
; APPLICANT: Gettitsen, Mary E.
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Sherwood, Steven
; APPLICANT: Smith, Victoria
; APPLICANT: Stewart, Timothy A.
; APPLICANT: Tumas, Daniel
; APPLICANT: Watanabe, Colin K
; APPLICANT: Wood, William
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; FILE REFERENCE: P3330R1C353
; CURRENT APPLICATION NUMBER: US/10/175,746
; CURRENT FILING DATE: 2002-06-19
; Prior Application removed - See File Wrapper or Palm
; NUMBER OF SEQ ID NOS: 550
; SEQ ID NO 222
; LENGTH: 290
; TYPE: PRT
; ORGANISM: Homo Sapien
US-10-175-746-222

Alignment Scores:
Pred. No.: 0.235 Length: 290
Score: 99.00 Matches: 49
Percent Similarity: 36.90% Conservative: 20
Best Local Similarity: 26.20% Mismatches: 48
Query Match: 3.69% Indels: 70
DB: Gaps: 9

US-10-020-441-1 (1-1477) x US-10-175-746-222 (1-290)
OY 91 TGTACAGSTTTCAGTCTGCTCAACGAGACGATCTCAGCTGCTGCTATTTGCTCTCA 150
Db 60 CysGlyGlySerLeuIleAlaGluGlnTrpValLeuThrAlaAlaHisCysPhe----- 77
OY 151 CCATTGCCAGTGATTCGGGTGAAGATGACATGCAAAACATTTGTTCTCATGTATCGA 210
Db 78 ----- 78
OY 211 TTGATTTACAGACAGCTGTTTCCGTGCGTGGCGTTTGTGTTGTATGACGTGGGTTT 270
Db 79 AsnThrSerGluThr----- 83
OY 271 GCATGATTTTAATGCTATATATCTTGGTCTTATTTCAAGTTTCAATTTCTCAGCTGAGG 330
Db 84 -----SerLeuTrpGlnValLeu----- 90
OY 331 AATGGCGAACCAAGGATCCATCACCACCGTCTGAGTTAAGGTGCACACGAGATAC 390
Db 91 ---GlyAlaArgGln-----LeuValGlnProGly--- 99
OY 391 ATGCCCTCTTGTATGTCGGCAGCAGAGAGA-----CCATTCGCACAG--- 435
Db 100 ---ProHisAlaMetTyrAlaArgValArgGlnValGluSerAsnProLeuTyrGlnGly 118
OY 436 ACACTCAGTGGATTTGATATGCAATTGTAATGCTGGCTCAATAGTGTCACTTACAGAGT 495
Db 119 ThrAlaSerSerAlaAspValAlaLeuValGluLeuGlnAlaProValProPheThrAsn 138
OY 496 GGAATCAGAGTATGATGCTGCGCACAGCATCGATATCCGCCACCTGAGACTGGTGT 555
Db 139 TyrIleLeuProValCysLeuProAspProSerValIlePheGluThrGlyMetCys 158
OY 556 TTCATTGTTGGTTATGAGAG---GATGATTAACGACCGTATCCGTCAACGTAAAGATGCT 612
Db 159 TrpValThrGlyTyrGlySerProSerGluGluAspLeuLeuProGluProArg----- 176
```



QY 271 GCATGATTTAAATGCTATATACCTGTTCTTATTTCAAGTTTCATTCTCACTGAGG 330  
|||:::|||||  
Db 84 -----SerLeuTyrGlnValLeu----- 90  
QY 331 AATGGCGACCAACAGGATTCATCCACCGCTGCGATTAAAGTGGCACCGAGATAC 390  
|||:::|||||  
Db 91 --GlyAlaArgGln-----LeuValGlnProGly--- 99  
QY 391 ATGCCCTCTTGATGTCGGCAGACAGAGAGA-----CCATGCCACAG--- 435  
|||:::|||||  
Db 100 ---ProHISalameTyrAlaArgValAlaGlnValGluSerAspProLeuTyrGlnGly 118  
QY 436 ACACCTGATGATTCGATATTCGATTTGATGCTGCTCAATGCTCAACTTACAGAGT 495  
|||:::|||||  
Db 119 ThrAlaSerSerAlaAspValAlaLeuValGluLeuGlnAlaProValProPheThrAsn 138  
QY 496 GGAATCAGAGTATCATCTCTGCCACAGCCATCGATATCCGCCACCTGGAACTGGTGT 555  
|||:::|||||  
Db 139 TyrIleLeuProValCysLeuProAspProSerValIlePheGluThrGlyMetAsnCys 158  
QY 556 TTCATTGTTGGTTATGAGAG---GATGATACGACCGTATCCGTACGTAAGATGT 612  
|||:::|||||  
Db 159 TrpValThrGlyTyrGlySerProSerGlnGluAspLeuLeuProGluProArg----- 176  
QY 613 GGAATATTGAAGAAAGTAGT 633  
|||:::|||||  
Db 177 ---IleLeuGlnIlybLeuAla 182  
RESULT 37  
US-10-137-865-222  
; Sequence 222, Application US/10137865  
; Publication No. US20030032155A1  
; GENERAL INFORMATION:  
; APPLICANT: Baker, Kevin P.  
; APPLICANT: Beresini, Maureen  
; APPLICANT: DeForge, Laura  
; APPLICANT: Desnoyers, Luc  
; APPLICANT: Filvaroff, Ellen  
; APPLICANT: Gao, Wei-Qiang  
; APPLICANT: Gerritsen, Mary E.  
; APPLICANT: Goddard, Audrey  
; APPLICANT: Godowski, Paul J.  
; APPLICANT: Gurney, Austin L.  
; APPLICANT: Sherwood, Steven  
; APPLICANT: Smith, Victoria  
; APPLICANT: Stewart, Timothy A.  
; APPLICANT: Tumas, Daniel  
; APPLICANT: Watanabe, Colin K  
; APPLICANT: Wood, William  
; APPLICANT: Zhang, Zemin  
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC  
; FILE REFERENCE: P3330R1C154  
; CURRENT APPLICATION NUMBER: US/10/137, 865  
; CURRENT FILING DATE: 2002-05-03  
; Prior Application removed - See Palm or File Wrapper  
; NUMBER OF SEQ ID NOS: 550  
; SEQ ID NO 222  
; LENGTH: 290  
; TYPE: PRT  
; ORGANISM: Homo Sapien  
US-10-137-865-222  
Alignment Scores:  
Pred. No.: 0.235 Length: 290  
Score: 99.00 Matches: 49  
Percent Similarity: 36.90% Conservative: 20  
Best Local Similarity: 26.20% Mismatches: 48  
Query Match: 3.69% Indels: 70  
DB: 9 Gaps: 9  
US-10-020-441-1 (1-1477) x US-10-137-865-222 (1-290)

QY 91 TGTACAGSTTCACTAGTCTCAACGAGACAGTACTACAGCTGTGATGTTGTCTCA 150  
|||:::|||||  
Db 60 CysGlyGlySerLeuIleAlaGluGlnIleTrpValLeuThrAlaIleCysPhe----- 77  
QY 151 CCATTGCCAGTATTCGGGTAAAGATGATGTAACAATTTGTTGCTCAATGTATGCA 210  
|||:::|||||  
Db 78 -----Arg 78  
QY 211 TTGATTTCACAGACAGTGTTCGCTGTCGTCGCTTGTGTTGTATGACAGTGTTT 270  
|||:::|||||  
Db 79 AsnThrSerGluThr----- 83  
QY 271 GCATGATTTAAATGCTATATACCTGTTCTTATTTCAAGTTTCATTCTCACTGAGG 330  
|||:::|||||  
Db 84 -----SerLeuTyrGlnValLeu----- 90  
QY 331 AATGGCGACCAACAGGATTCATCCACCGCTGCGATTAAAGTGGCACCGAGATAC 390  
|||:::|||||  
Db 91 --GlyAlaArgGln-----LeuValGlnProGly--- 99  
QY 391 ATGCCCTCTTGATGTCGGCAGACAGAGAGA-----CCATGCCACAG--- 435  
|||:::|||||  
Db 100 ---ProHISalameTyrAlaArgValAlaGlnValGluSerAspProLeuTyrGlnGly 118  
QY 436 ACACCTGATGATTCGATATTCGATTTGATGCTGCTCAATGCTCAACTTACAGAGT 495  
|||:::|||||  
Db 119 ThrAlaSerSerAlaAspValAlaLeuValGluLeuGlnAlaProValProPheThrAsn 138  
QY 496 GGAATCAGAGTATCATCTCTGCCACAGCCATCGATATCCGCCACCTGGAACTGGTGT 555  
|||:::|||||  
Db 139 TyrIleLeuProValCysLeuProAspProSerValIlePheGluThrGlyMetAsnCys 158  
QY 556 TTCATTGTTGGTTATGAGAG---GATGATACGACCGTATCCGTACGTAAGATGT 612  
|||:::|||||  
Db 159 TrpValThrGlyTyrGlySerProSerGlnGluAspLeuLeuProGluProArg----- 176  
QY 613 GGAATATTGAAGAAAGTAGT 633  
|||:::|||||  
Db 177 ---IleLeuGlnIlybLeuAla 182  
RESULT 38  
US-10-140-474-222  
; Sequence 222, Application US/10140474  
; Publication No. US20030032155A1  
; GENERAL INFORMATION:  
; APPLICANT: Baker, Kevin P.  
; APPLICANT: Beresini, Maureen  
; APPLICANT: DeForge, Laura  
; APPLICANT: Desnoyers, Luc  
; APPLICANT: Filvaroff, Ellen  
; APPLICANT: Gao, Wei-Qiang  
; APPLICANT: Gerritsen, Mary E.  
; APPLICANT: Goddard, Audrey  
; APPLICANT: Godowski, Paul J.  
; APPLICANT: Gurney, Austin L.  
; APPLICANT: Sherwood, Steven  
; APPLICANT: Smith, Victoria  
; APPLICANT: Stewart, Timothy A.  
; APPLICANT: Tumas, Daniel  
; APPLICANT: Watanabe, Colin K  
; APPLICANT: Wood, William  
; APPLICANT: Zhang, Zemin  
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC  
; FILE REFERENCE: P3330R1C162  
; CURRENT APPLICATION NUMBER: US/10/140, 474  
; CURRENT FILING DATE: 2002-05-06  
; Prior Application removed - See Palm or File Wrapper  
; NUMBER OF SEQ ID NOS: 550  
; SEQ ID NO 222  
; LENGTH: 290  
; TYPE: PRT

ORGANISM: Homo Sapien  
US-10-140-474-222

## Alignment Scores:

Pred. No.:	0.235	Length:	290
Score:	99.00	Matches:	49
Percent Similarity:	36.90%	Conservative:	20
Best Local Similarity:	26.20%	Mismatches:	48
Query Match:	3.69%	Indels:	70
DB:	9	Gaps:	9

US-10-020-441-1 (1-1477) x US-10-140-474-222 (1-290)

```
QY 91 TGTACAGSTTCACTAGTCTCAACAGACAGTACTCAGAGTGTGCTATGTTGCTCA 150
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 60 Cysgllyglserleuillaalaglulntrpvalleuthralaiahscyphe----- 77
QY 151 CCATTGCCAGTATCGGGTAAGAGATGACTGAACAACATTTGTTGCTCATGTATGCA 210
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 78 -----Arg 78
QY 211 TTGATTTCAGACAGAGTGTTCGGTGGCGTGTGCTTTGTGTTGTATGACAGTGGGTTT 270
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 79 Asnhrsergluthr-----Arg 83
QY 271 GCATGATTTTAATGCTATATACTTGGCTTATTTCAGTTTCATTTCTACACTGAGG 330
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 84 -----Serleuylrglnvalleu----- 90
QY 331 AATGCCGCAACAAGGATCCATCCATCCACCGTCTGAGTTAAGTGGCACCAGGATAC 390
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 91 --GlyalaArgGln-----LeuValGlnProGly--- 99
QY 391 ATGCCCTCTTGATGTCGGCAGACAGAGAGA-----CCATGCCACAG--- 435
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 100 ---ProhlsalameTyrAlaArgValAlaArgGlnValGluSerAsnProleuylrglnGly 118
QY 436 ACACCTCAGTGAATTCGATTTGCAATTTGATGCTGGCTCAATGCTCAACTTACAGAGT 495
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 119 ThrAlaSerSerAlaAspValAlaLeuValGluLeuGlnAlaProValProPheThrAsn 138
QY 496 GGAATCAGAGTATCATGCTGCCACAGCCATGATATCCGCCACCTGGAACTGGTGT 555
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 139 TyrIleLeuProValCysLeuProAspProSerValIlePheGlnThrGlyMetAsnCys 158
QY 556 TTCATTGTTGGTTATGAGAG---GATGATTAAGACCGGATCCGTCAGTAAAGATGT 612
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 159 TrpValThrGlyTyrGlySerProSerGluGluAspLeuProGluProArg----- 176
QY 613 GGAATATTGAAGAAAGTGAAGT 633
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 177 ---IleLeuGlnIlylsleuAla 182

RESULT 39
US-10-142-431-222
; Sequence 222, Application US/10142431
; Publication No. US20030036179A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Beresini, Maureen
; APPLICANT: DeForge, Laura
; APPLICANT: Desnoyers, Luc
; APPLICANT: Filvaroff, Ellen
; APPLICANT: Gao, Wei-Qiang
; APPLICANT: Gerritsen, Mary E.
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Sherwood, Steven
; APPLICANT: Smith, Victoria
; APPLICANT: Stewart, Timothy A.
; APPLICANT: Tumas, Daniel
; APPLICANT: Watanabe, Colin K
```

```
; APPLICANT: Wood, William
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; FILE REFERENCE: P330R1C251
; CURRENT APPLICATION NUMBER: US/10/142,431
; CURRENT FILING DATE: 2002-05-10
; Prior Application removed - See file wrapper or Palm
; NUMBER OF SEQ ID NOS: 550
; SEQ ID NO 222
; LENGTH: 290
; TYPE: PRT
; ORGANISM: Homo Sapien
US-10-142-431-222
```

## Alignment Scores:

Pred. No.:	0.235	Length:	290
Score:	99.00	Matches:	49
Percent Similarity:	36.90%	Conservative:	20
Best Local Similarity:	26.20%	Mismatches:	48
Query Match:	3.69%	Indels:	70
DB:	9	Gaps:	9

US-10-020-441-1 (1-1477) x US-10-142-431-222 (1-290)

```
QY 91 TGTACAGSTTCACTAGTCTCAACAGACAGTACTCAGAGTGTGCTATGTTGCTCA 150
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 60 Cysgllyglserleuillaalaglulntrpvalleuthralaiahscyphe----- 77
QY 151 CCATTGCCAGTATCGGGTAAGAGATGACTGAACAACATTTGTTGCTCATGTATGCA 210
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 78 -----Arg 78
QY 211 TTGATTTCAGACAGAGTGTTCGGTGGCGTGTGCTTTGTGTTGTATGACAGTGGGTTT 270
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 79 Asnhrsergluthr-----Arg 83
QY 271 GCATGATTTTAATGCTATATACTTGGCTTATTTCAGTTTCATTTCTACACTGAGG 330
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 84 -----Serleuylrglnvalleu----- 90
QY 331 AATGCCGCAACAAGGATCCATCCATCCACCGTCTGAGTTAAGTGGCACCAGGATAC 390
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 91 --GlyalaArgGln-----LeuValGlnProGly--- 99
QY 391 ATGCCCTCTTGATGTCGGCAGACAGAGAGA-----CCATGCCACAG--- 435
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 100 ---ProhlsalameTyrAlaArgValAlaArgGlnValGluSerAsnProleuylrglnGly 118
QY 436 ACACCTCAGTGAATTCGATTTGCAATTTGATGCTGGCTCAATGCTCAACTTACAGAGT 495
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 119 ThrAlaSerSerAlaAspValAlaLeuValGluLeuGlnAlaProValProPheThrAsn 138
QY 496 GGAATCAGAGTATCATGCTGCCACAGCCATGATATCCGCCACCTGGAACTGGTGT 555
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 139 TyrIleLeuProValCysLeuProAspProSerValIlePheGlnThrGlyMetAsnCys 158
QY 556 TTCATTGTTGGTTATGAGAG---GATGATTAAGACCGGATCCGTCAGTAAAGATGT 612
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 159 TrpValThrGlyTyrGlySerProSerGluGluAspLeuProGluProArg----- 176
QY 613 GGAATATTGAAGAAAGTGAAGT 633
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 177 ---IleLeuGlnIlylsleuAla 182

RESULT 40
US-10-143-114-222
; Sequence 222, Application US/10143114
; Publication No. US20030036180A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Beresini, Maureen
; APPLICANT: DeForge, Laura
```

APPLICANT: Desnoyers, Luc  
APPLICANT: Filvaroff, Ellen  
APPLICANT: Gao, Wei-Qiang  
APPLICANT: Gerritsen, Mary E.  
APPLICANT: Goddard, Audrey  
APPLICANT: Godowski, Paul J.  
APPLICANT: Gurney, Austin L.  
APPLICANT: Sherwood, Steven  
APPLICANT: Smith, Victoria  
APPLICANT: Stewart, Timothy A.  
APPLICANT: Tumas, Daniel  
APPLICANT: Watanabe, Colin K  
APPLICANT: Wood, William  
APPLICANT: Zhang, Zemin  
TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC  
FILE REFERENCE: P330R1C211  
CURRENT APPLICATION NUMBER: US/10/143,114  
CURRENT FILING DATE: 2002-05-09  
Prior Application removed - See Palm or File Wrapper  
NUMBER OF SEQ ID NOS: 550  
SEQ ID NO 222  
LENGTH: 290  
TYPE: PRT  
ORGANISM: Homo Sapien  
US-10-143-114-222

Alignment Scores:  
Pred. No.: 0.235 Length: 290  
Score: 99.00 Matches: 49  
Percent Similarity: 36.90% Conservative: 20  
Best Local Similarity: 26.20% Mismatches: 48  
Query Match: 3.69% Indels: 70  
DB: 9 Gaps: 9

US-10-020-441-1 (1-1477) x US-10-143-114-222 (1-290)

QY 91 TGTACAGSTTCACTAGTCTCAACGACGACTCAAGCTGATGTTGTTCTCA 150  
DB 60 CysGlyGlySerLeuIleAlaGlnIntrPvalLeuThrAlaAlaHisCysPhe----- 77  
QY 151 CCATTGCCAGTATGGGTAAAGATGACTGAACAACATTGTTGCTCAATGTAATCGA 210  
DB 78 -----Arg 78  
QY 211 TTGATTTACAGACAGCTGTTGCGTGCCTGCTTGTGTTGTATGACAGTGGTTT 270  
DB 79 AsnThrSerGluThr-----Arg 83  
QY 271 GCATGATTTTAATGCTATATACTGCTTATTTTCAAGTTTCATTCTCACACTGAGG 330  
DB 84 -----SerLeuTyrGlnValLeuLeu----- 90  
QY 331 AATGGCGACCAACAGGATCCATCCACCAACCGTGTGAGTTAAGTGGCACACAGATAC 390  
DB 91 --GlyAlaArgGln-----LeuValGlnProGly--- 99  
QY 391 ATGCCCTTGTATGTCGCGACGACGAGAGAGA-----CCAATGCGACAG--- 435  
DB 100 --ProHisAlaMetTyrAlaArgValArgGlnValGlnSerAsnProLeuTyrGlnGly 118  
QY 436 ACACTAGTGAATTCGATTTGCAATTTGAATGCTGGCTCAATAGTGTCACTTACAGAGT 495  
DB 119 ThrAlaSerSerAlaAspValAlaLeuValGlnLeuGlnAlaProValProPheThrAsn 138  
QY 496 GGAATCAGAGTCAAGTCTGCCACAGCAGCATGATATCCCGCACTGGAACCTGAGT 555  
DB 139 TyrIleLeuProValCysLeuProAspProSerValIlePheGlnThrGlyMetAsnCys 158  
QY 556 TTCATGTTGTTATGGAAG---GATGATAACGACCGTATCCGTCAAGTAAAGATGT 612  
DB 159 TrpValThrGlyTyrPglySerProSerGlnGlnAlaLeuProGlnProArg----- 176

QY 613 GGAATATTGAAGAAAGTGAGT 633  
DB 177 ---IleLeuGlnIysLeuAla 182

Search completed: April 9, 2003, 13:56:07  
Job time : 84.5 secs

**THIS PAGE BLANK (USPTO)**



GenCore version 5.1.4 p5 4578  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - protein search, using frame\_plus\_n2p model

Run on: April 9, 2003, 13:25:40 ; Search time 127.5 Seconds  
(without alignments)  
4773.829 Million cell updates/sec

Title: US-10-020-441-1  
Perfect score: 2686  
Sequence: 1 gtgtcaaccgcgtgtacacg.....ttgtacgtccaatcttga 1477

Scoring table: BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 1343160

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Command line parameters:

-MODEL=frame+n2p model -DEV=xl1p  
-Q=/cgn2\_1/USPTO.spool/US10020441/runat\_09042003\_121211\_1340/app\_query.fasta\_1.1671  
-DB=SPTREMBL\_21 -OFMT=faetan -SUFFIX=rspt -MINMATCH=0.1 -LOOPEXT=0 -LOOPEXT=0  
-UNITS=bits -START=1 -END=1 -MATRIX=biosum62 -TRANS=human40.cdi -LIST=45  
-DOCALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=40 -MODE=LOCAL  
-OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000  
-USRR=US10020441.CCN 1.1 172 @runat\_09042003\_121211\_1340 -NCPU=6 -ICPU=3  
-NO\_XLPHY -NO\_MMALP -LARGEOVERY -NEG\_SCORES=0 -WAIT -LONGLOG -DEV-TIMEOUT=120  
-WARN\_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -Fgapop=6 -Fgapext=7  
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

SPTREMBL\_21: \*  
1: sp\_archaea: \*  
2: sp\_bacteria: \*  
3: sp\_fungi: \*  
4: sp\_human: \*  
5: sp\_invertebrate: \*  
6: sp\_mammal: \*  
7: sp\_mhc: \*  
8: sp\_organelle: \*  
9: sp\_phage: \*  
10: sp\_plant: \*  
11: sp\_rodent: \*  
12: sp\_virus: \*  
13: sp\_vertebrate: \*  
14: sp\_unclassified: \*  
15: sp\_tvirus: \*  
16: sp\_bacteriopl: \*  
17: sp\_archaeap: \*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1094	40.7	274	5 Q26552	Q26552 schistosoma

2	825.5	30.7	257	5	Q26553	Q26553 schistosoma
C 3	152	5.7	5532	2	Q9S0R4	Q9S0R4 streptomyc
4	144	5.4	421	5	Q9SR66	Q9SR66 drosophila
C 5	139	5.2	623	5	P90533	P90533 dictyosteli
C 6	138	5.2	6239	2	Q9S0R7	Q9S0R7 streptomyc
7	137	5.1	978	5	P91777	P91777 pacifastacu
8	136	5.1	307	4	Q9H9U3	Q9H9U3 homo sapien
C 9	133.5	5.0	1217	4	Q9UKW9	Q9UKW9 homo sapien
C 10	132.5	5.0	957	4	Q14651	Q14651 homo sapien
C 11	128.5	4.8	3972	2	Q9S0R8	Q9S0R8 streptomyc
C 12	126.5	4.7	1795	5	Q76894	Q76894 drosophila
C 13	124.5	4.7	1171	5	Q9VHC2	Q9VHC2 drosophila
C 14	124.5	4.7	3570	4	Q99552	Q99552 homo sapien
C 15	124	4.6	237	6	Q29464	Q29464 bos taurus
16	123.5	4.6	272	5	Q18459	Q18459 heterodera
17	123	4.6	233	4	Q96R27	Q96R27 homo sapien
18	123	4.6	275	4	Q96R26	Q96R26 homo sapien
C 19	122.5	4.6	3626	2	Q9F779	Q9F779 streptomyc
C 20	121.5	4.6	505	4	Q14395	Q14395 homo sapien
21	121	4.5	281	5	Q76898	Q76898 drosophila
22	121	4.5	365	5	Q97366	Q97366 holotrichia
C 23	121	4.5	477	4	Q14887	Q14887 homo sapien
C 24	121	4.5	957	4	Q9UKR0	Q9UKR0 homo sapien
C 25	120.5	4.5	559	5	Q9U3Y8	Q9U3Y8 plasmodium
C 26	120.5	4.5	1349	4	Q8WQ4	Q8WQ4 homo sapien
27	120	4.5	242	4	Q96R25	Q96R25 homo sapien
C 28	120	4.5	740	5	Q8T138	Q8T138 dictyosteli
29	118.5	4.4	974	13	Q90WD8	Q90WD8 bufo japoni
30	118	4.4	270	4	Q96QU8	Q96QU8 homo sapien
31	117	4.4	166	4	Q9UQ18	Q9UQ18 homo sapien
C 32	116.5	4.4	858	4	Q9H8D9	Q9H8D9 homo sapien
C 33	116.5	4.4	1492	5	Q8SSU1	Q8SSU1 dictyosteli
C 34	115	4.3	604	5	Q26021	Q26021 plasmodium
C 35	115	4.3	1461	5	Q8ST04	Q8ST04 dictyosteli
C 36	115	4.3	1976	5	Q9V9I8	Q9V9I8 drosophila
37	114.5	4.3	257	6	Q19023	Q19023 mecacaca mlla
38	114	4.2	273	6	Q9XSM1	Q9XSM1 ovis aries
C 39	114	4.2	1472	5	Q9V358	Q9V358 drosophila
C 40	113.5	4.2	298	5	Q8T4N4	Q8T4N4 ribiccephal
C 41	113.5	4.3	746	5	Q9V515	Q9V515 drosophila
C 42	113	4.2	255	3	Q9Y7A9	Q9Y7A9 metarhizium
C 43	113	4.2	604	13	Q42472	Q42472 xenopus lae
C 44	112.5	4.2	273	12	P87519	P87519 bovine herp
C 45	112.5	4.2	961	3	Q92223	Q92223 emericella

## ALIGNMENTS

### RESULT 1

ID	Q26552	PRELIMINARY;	PRT;	274 AA.
AC	Q26552;			
DT	01-NOV-1996 (TREMBLrel. 01, Created)			
DT	01-NOV-1996 (TREMBLrel. 01, Last sequence update)			
DT	01-MAR-2002 (TREMBLrel. 20, Last annotation update)			
DE	Elastase.			
OS	Schistosoma mansoni (Blood fluke).			
OC	Eukaryota; Metazoa; Platyhelminthes; Trematoda; Digenea; Strigeidida;			
OC	Schistosomatidae; Schistosomatidae; Schistosoma.			
OX	NCBI_TaxID=6183;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=PUERTO RICO;			
RX	MEDLINE=96362066; PubMed=8720180;			
RA	Pierrot C., Capron A., Khalife J.;			
RT	"Cloning and characterization of two genes encoding Schistosoma			
RT	mansoni elastase.";			
RL	Mol. Biochem. Parasitol. 75:113-117(1995).			
CC	-1 SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE			
CC	TRYPSIN FAMILY.			
DR	EMBL; U31768; AAC46967.1; -.			
DR	HSSP; P20231; IAAO.			
DR	MEROFS; S01.144; -.			

DR InterPro: IPR001314; Chymotrypsin.  
 DR InterPro: IPR001254; Ser\_protease\_Try.  
 DR Pfam: PF00089; trypsin; 1.  
 DR PRINTS: PR00722; CHYMOTRYPSIN.  
 DR SMART: SM00020; Tryp\_Spc; 1.  
 DR PROSITE: PS0240; TRYPSIN\_DOM; 1.  
 DR PROSITE: PS00134; TRYPSIN\_HIS; UNKNOWN\_1.  
 DR HydroLase: Serine protease.  
 KW SEQUENCE 274 AA; 29586 MW; 53F21914FD3DA450 CRC64;

## Alignment Scores:

Pred. No.: 2.09e-105 Length: 274  
 Score: 1094.00 Matches: 245  
 Percent Similarity: 50.00% Conservative: 1  
 Best Local Similarity: 49.80% Mismatches: 7  
 Query Match: 40.73% Indels: 239  
 DB: Gaps: 4

US-10-020-441-1 (1-1477) x Q26552 (1-274)

QY 1 GTGTCAACCTGGTGTATACGTAGTGGTGAACCTGTGCAACACCGCACTGAATTCGGCTTC 60  
 DB 22 ValSerThrTrpLeuIleArgSerGlyIuproValGlnHisArgThrGluheProPhe 41  
 QY 61 ATGCAATCTTAACGACAGAGAAACAATGTGTACAGSTTCACTAGTCTCAACGAGAGA 120  
 DB 42 IleAlaPheLeuThrThrGluArgThrMetCysThrGlySerLeuValSerThrArgAla 61  
 QY 121 GATCTCACAGCGTGTGATGTTGGTCTGACCATGGCCAGTATTCGGGTAAAGATGCA 180  
 DB 62 ValLeuThrAlaGlyHisCysValCysSerProLeuProValIleArgVal 78  
 QY 181 CTGAACAACATGTGTGCTCAATGATTCGATTGATTCACAGACAGTGTGCGTGTGCG 240  
 DB 78 78  
 QY 241 TGGCTTTGTGTTGTGTATGATGAGTGGGTTTGCAATGATTTAATGTCATATACCTTGTC 300  
 DB 79 79  
 QY 301 TTATTTTCAGGTTTCATTTTCTGCACTGAGGAATGGGACCAACAGGATCCATACCCA 360  
 DB 81 LeuPheGlnValSerPheLeuThrLeuArgAsnGlyAspGlnGlnIleHisShiSgln 100  
 QY 361 CCGTGTGAGGTAAAGGTGGCACCAAGATACATGCCCTCTGTATGTCGGACGACAGAG 420  
 DB 101 ProSerGlyValAllyValAlaProGlyTyrMetProSerCysMetSerAlaArgArgGly 120  
 QY 421 AGACCAATGCAACAGACTCACTGAGTGAATTCGATTTGCAATTTGTAATGCTGCTCAATG 480  
 DB 121 ArgProIleAlaGlnThrLeuSerGlyPheAspIleAlaIleValMetLeuAlaGlnMet 140  
 QY 481 GTCACTTTCAGAGTGAATCAGAGTCAAGTCACTGCCACAGCATCGGATATCCGCCA 540  
 DB 141 ValAsnLeuGlnSerGlyIleThrValIleSerLeuProGlnAlaSerAspIleProThr 160  
 QY 541 CCTGGAACGTGTTGTTTCATTTGTTGTTATGGAAGGATGATTAACAACCGTATCCGTCA 600  
 DB 161 ProGlyThrGlyValAlaPheIleValGlyTyrGlyArgAspAsnAspArgAspProSer 180  
 QY 601 CGTAAGAATGTGGGAATATTAAGAAAGTGAAGTGTGGTGAATAAACAATGACTCAG 660  
 DB 181 ArgGlyAsnGlyGlyIleLeuIlySlys 189  
 QY 661 TGTGTAGTCAAGTCAAGTCAAGTATATCTGTGTGTCTGTATCTGTTGTGTCTGTCTGTC 720  
 DB 189 189  
 QY 721 TGTCTACTGATCCGGTGTGTGATTTGTCAGAGCCTTGATATAACAACATGTTTGA 780  
 DB 189 189  
 QY 781 TGACTTTTGACAGTTCAAGTAGACAGAGATTTTCATCTCGGTCATGTGTGGTGAGGT 840

DB 189 189  
 QY 841 GAGTGACGTATGTGAGTGAGTGAGTTAGAGTAGAATTGATGGATGGAATGTGATGTGATG 900  
 DB 189 189  
 QY 901 GGATGATTGAGACCACTTGAGAGAGAGAGAAGCTCATGAAATATCTTATGCAAAACATGGA 960  
 DB 190 190  
 QY 961 GTGTGTGTGTGATCAATGAAGTACAGGGGTCAATGTGTTAGATATGTGTTTGAGAGAGTG 1020  
 DB 191 191  
 QY 1021 AGATGAGAGTGACTTGTATCGTCGAATATATAGTACATGTGATGTATGTGACTATTGT 1080  
 DB 191 191  
 QY 1081 TGTGTGGGTAGTGTAAAGGTGAGATATTTGTGCCAGTGTATTTTGAATTCAGTTGTG 1140  
 DB 192 193  
 QY 1141 TGTGTTGTTGTTGTTCCGTGAGTTGTTGTTGTTCTTTCATCTGACTGTACGTTGTTGACT 1200  
 DB 194 194  
 QY 1201 GATCTGTTGTGTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTT 1260  
 DB 195 195  
 QY 1261 ACATGCAACCAATGCAATCTATATGTGTGAAGACAGTCAAGATTTGCGACAGTTACC 1320  
 DB 203 GHisAlaThrAsnGlyAsnProIleCysValIlyAsnGlyGlnAsnPheGlyGlnLeuPr 223  
 QY 1321 CGCTCAGGTGACATGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTT 1380  
 DB 223 AlaProGlyAspSerGlyGlyProLeuLeuProSerProGlnGlyProValLeuGlyVa 243  
 QY 1381 CGTATCACATGGTGTGACACCTGCCCAACCTTCCGATATCATTTGTCAGTATGCCAGT 1440  
 DB 243 ValSerHisGlyValIleThrLeuProAsnLeuProAspIleIleValGlyTyrAlaSerVa 263  
 QY 1441 GGTAGAATGTTGATTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTT 1474  
 DB 263 AlaArgMetLeuAspPheValArgSerAsnIle 274

RESULT 2  
 Q26553 PRELIMINARY; PRT; 257 AA.  
 AC Q26553;  
 DT 01-NOV-1996 (TrEMBLrel. 01, Created)  
 DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)  
 DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)  
 DE Elastase.  
 OS Schistosoma mansoni (Blood fluke).  
 OC Eukaryota; Metazoa; Platyhelminthes; Trematoda; Digenea; Strigeidida;  
 OC Schistosomatidae; Schistosomatidae; Schistosoma.  
 OX NCBI\_TaxID=6183;  
 RN (1)  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=PUERTO RICO;  
 RX MEDLINE=96362066; PubMed=8720180;  
 RA PIERROT C., CAPRON A., KHALIFE J.;  
 RT "Cloning and characterization of two genes encoding Schistosoma  
 mansoni elastase.";  
 RL MOL. Biochem. Parasitol. 75:113-117(1995).  
 DR HSSP; P00766; IGCD.  
 DR InterPro: IPR001254; Ser\_protease\_Try.  
 DR Pfam; PF00089; trypsin; 1.  
 DR SMART; SM00020; Tryp\_Spc; 1.  
 DR PROSITE; PS0240; TRYPSIN\_DOM; 1.

DR PROSITE, PS00134; TRYPSIN\_HIS; UNKNOWN\_1.  
 KW Hydrolyase; Serine protease.  
 SQ SEQUENCE 257 AA; 27971 MM; 00F8872799F9DD64 CRC64;

## Alignment Scores:

Pred. No.:	2,476-77	Length:	257
Score:	825.50	Matches:	190
Percent Similarity:	42.07%	Conservative:	17
Best Local Similarity:	38.62%	Mismatches:	29
Query Match:	30.73%	Indels:	256
DB:	5	Gaps:	3

US-10-020-441-1 (1-1477) x 026553 (1-257)

```

QY 1 GTGTCAACCTGGTGTATAGTAGTGGTGAACCTGTGCAACACCGCATGATCCGGTTC 60
   |||||
DB 22 VALSERHTTPLEHILIEKRSERAPGILNPROVALGILYHISHTHRLINPHEPROPE 41
   |||||
QY 61 ATCGCATCTTAAACAGACAGAGAAACATGTGTACAGSTTCACTAGTCTCAACGAGACA 120
   |||||
DB 42 ILEALTYRILEALSERLYSERYSERMETCYSTHNGLYSERLEUVALSERHTHRAVAL 61
   |||||
QY 121 GTACTCACAGCTGCTCATTTGTGTCTCAACCATTCGCCGTGATTCGGGTAAAGATGCA 180
   |||||
DB 62 VALLEUTHRLAAGLYHISCYVALCYSPROPMETPROVALILELYVAL 78
   |||||
QY 181 CTGAACAACATGTGTCTCAATGTAATGATTTGACAGACAGTGTTCGGTGTGCG 240
   |||||
DB 78 ----- 78
   |||||
QY 241 TGCGTTGTGTGTGTGTATGACAGTGGGTTTGACATGATTTATGTCATATACCTGTGTC 300
   |||||
DB 78 ----- 78
   |||||
QY 301 TTATTTCAAGTTTCACTTTCTCACTAGAGAAATGGGACCAACAGCATTCATCAACCA 360
   |||||
DB 79 -----GILNVALTHRLPHEUTHRLLEUHRGASNGLYASPGLNGILYILEHISHISGLN 96
   |||||
QY 361 CGGTCTGAGGTGAAGTGGACACAGGATACATGCCCTCTGTATGTCGGGACACAGAGG 420
   |||||
DB 97 PROSERGLVALALYVALALAPROGLUTRYMETPROSERCYTHRLALSERHGLINAG 116
   |||||
QY 421 AGACCAATCGACACAGACACTCAGTGGATTCGATTTGCAATGTGTAATGCTGCTCAATG 480
   |||||
DB 117 ARGARGILEARGGLNTHRLSEUSERGLYPHEALILEALATHRLMETLEUHLAGLINMET 136
   |||||
QY 481 GTCAACTTACAGAGTGAATCAGAGTATCAGTCTGCCACAGCCATCGGATATCCCGCCA 540
   |||||
DB 137 VALASNLEUGINSERGLYLEARGVALILEGLYLEUPROGLNALASERASPILEPROTHR 156
   |||||
QY 541 CCTGGAACCTGGTGTTCATTTGTTATGTAAGAGGATGAAGGATGAACCGGTATCCGTCA 600
   |||||
DB 157 PROGLYTHRAPVALPHELEVALGLYTYGLYARGASAPASAPASAPRTRGSPROSER 176
   |||||
QY 601 CGTGAAGATGTGAATATTGGAAGAAAGTGAAGTGTGGTGAATAAACACATGACTCAG 660
   |||||
DB 177 ARGARGILAEGLYGLY----- 181
   |||||
QY 661 TCGTCAGTCAAGTGTCACTCACTATTTCTGTGTGTCTGTATCTGTTGTGTCTGTCTGTC 720
   |||||
DB 181 ----- 181
   |||||
QY 721 TGTCTACCTGATCCGGTGTGTGTATTGTGTGTCAGAGCCTGTGATATAACAACACTGTGTTGA 780
   |||||
DB 181 ----- 181
   |||||
QY 781 TGACTTTGTGACAGTTCAGTACAGAGATGATTTCCATCTCGGTCAATGTGTTGTGAGGT 840
   |||||
DB 181 ----- 181
   |||||
QY 841 GAGTGACGTGATGTGAGGTGAGTGAAGTGAATGGATGGATGGATGGATGGATGGATG 900
   |||||
DB 181 ----- 181

```

```

QY 901 GGATGATTGAGACACTTGGAGAGAGAGAAAGCTCATGAATAATCTATGCAAAACGATGANA 960
   |||||
DB 181 ----- 181
   |||||
QY 961 GTGTGTGTGTATCATGAAGTAAAGGGGTCATATGTGTTGAGTATGTGTTGGAGAGTGTG 1020
   |||||
DB 181 ----- 181
   |||||
QY 1021 AGATGAGAGTACTTGTATCGTCAATATATAGTGCACATGTGATTTGATGTGACTATTGT 1080
   |||||
DB 181 ----- 181
   |||||
QY 1081 TGTGTGGTAGTGTAAAGGTGATATTGTCACAGTTGATATTTTTCGAATTCACCTGTG 1140
   |||||
DB 181 ----- 181
   |||||
QY 1141 TGTGTTGTTGTTCCGTTGTGTGTTGTTGTTCTTCATCTGACTGTACGTTGTTGTA 1200
   |||||
DB 181 ----- 181
   |||||
QY 1201 GTACTGTGTTGTGTGTTGTTGTTGTTGTCACACAGGCTGACGACATATATGGAATGCGG 1260
   |||||
DB 182 -----Gly-ArgAlaThrValThrGlnCysArg 190
   |||||
QY 1261 ACATGACCAATGGCAATCCATATGTGTGAAAGACAGGTGAGATTTCCGACAGTTACC 1320
   |||||
DB 190 GHISGLIUTHRLHISVALANPPOILIECYVALIYHISGLYPROHNSERGLYGLINLEU 210
   |||||
QY 1321 CGCTTCAGGTGACAGTGTGAGACCTCTCTCCCATCCCTTCAAGGTGACGATCTGGTGT 1380
   |||||
DB 210 UGLYPROGLYAPSERGLYGLYPROLEUENUPROSERPROGLINGLYPROVALILEGLYVA 230
   |||||
QY 1381 CGTATCACATGTGTGTCAACACTGCCAACTCCGATATCATGTGCGATATGCCAGTGT 1440
   |||||
DB 230 IVALSERHISGLYVALTHRLLEUHRGASNGLYASPGLNGILYILEHISHISGLN 246
   |||||
QY 1441 GGCTGAGATGTGTGATTTGTTGACGCTCCCATATT 1474
   |||||
DB 246 IVALSERHISGLYVALTHRLLEUHRGASNGLYASPGLNGILYILEHISHISGLN 257
   |||||

```

## RESULT 3

Q9S0R4 PRELIMINARY; PRT; 5532 AA.  
 ID Q9S0R4  
 AC Q9S0R4;  
 DT 01-MAY-2000 (TRENDS) 13, Created)  
 DT 01-MAY-2000 (TRENDS) 13, Last sequence update)  
 DT 01-MAR-2002 (TRENDS) 20, Last annotation update)  
 DE Type 1 polyketide synthase AVES 3.  
 GN AVEA3.  
 OS Streptomyces avermitilis.  
 OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;  
 OC Actinomycetales; Streptomycinae; Streptomycetaceae; Streptomyces.  
 OX NCBI\_TaxID=33903;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=99380548; PubMed=10449723;  
 RA Ikeda H., Nomomiya T., Usami M., Ohra T., Omura S.;  
 RT "Organization of the biosynthetic gene cluster for the polyketide  
 anhydromycin macroide avermectin in Streptomyces avermitilis";  
 RL Proc. Natl. Acad. Sci. U.S.A. 96:9509-9514(1999).  
 DR EMBL: AB032367; BAA84478.1;  
 DR InterPro: IPR001227; Ac transferase.  
 DR InterPro: IPR000794; Ketoacyl-synt.  
 DR InterPro: IPR003880; Pantone attach.  
 DR InterPro: IPR000834; Zn carbonyl.  
 DR Pfam: PF00698; Acyl transferase; 3.  
 DR Pfam: PF00109; ketoacyl-synt; 3.  
 DR Pfam: PF02801; ketoacyl-synt; C; 3.  
 DR Pfam: PF00550; PP-binding; 3.  
 DR PROSITE, PS00075; ACP DOMAIN; 3.  
 DR PROSITE, PS00606; B KETOACYL SYNTHASE; 3.  
 DR PROSITE, PS00133; CARBOXYPEPT\_ZN\_2; UNKNOWN\_1.



```

US-10-020-441-1 (1-1477) x Q95RS6 (1-421)
QY 49 GAATCCGTCATCGCATCTTAAAGACAGAGAACATG-----TGACAGST 99
    |||||
Db 175 GluthePOTPrthValAlaleuLeuHsSerGlyAsnLeuSerThrPheCysAlaGly 194
    |||||
QY 100 TCACAGTCTCAACAGAGCACTACTCAACAGTGTGATCTGTGTTGCTCAACATTGCGCA 159
    |||||
Db 195 SerLeuIleHsLysGlnValIleuThrAlaIleHsCysValGlySer----- 211
    |||||
QY 160 GTGATTCGGTGAAGATGCACTGAACAACATTTGGCTCAATGTAATGATGATTTCGA 219
    |||||
Db 211 ----- 211
QY 220 CAGACAGTGTTCGTCGTGCGTGGTGTGTTGTGTAAGCATGGTTGTCATAGATT 279
    |||||
Db 211 ----- 211
QY 280 TTAATGTCTAATACCTGTCATTTATTCAGGTTTCATTTTCACACTGAGAAAGCGCAC 339
    |||||
Db 212 -----LeuArgThrGlySerPheThrValArgAlaGlyGlu 223
    |||||
QY 340 -----CAACAGGCATCCATCCACCAACCGTCTGAGTTAAGTGCGCA 381
    |||||
Db 224 TrpAspThrGlnThrMetLysGlnArgLeuProTyrGlnGluArgSerValGlnThrVal 243
    |||||
QY 382 -----CCAGATACATGCCCTTGTATGTGGCAGCAGCAGAGGAGCAACATCGCA 432
    |||||
Db 244 IleLeuHsPheProAspPyr-----AspArgArgSerIleAla 255
    |||||
QY 433 CAGACACTCAGTGTGATTCGATTCGAATTTGTAATGCTGGCTCAATAGTCAACTTACAG 492
    |||||
Db 256 -----TyrAspPheAlaLeuValIleLeuSerGlnProValThrLeuAsp 270
    |||||
QY 493 AGTGAATCAGAGTATCATGTTGTCGACAGCCATCGATATCCCGCACCTGGAATCGGT 552
    |||||
Db 271 AspHisIleAsnValIleCysLeuProGlnIleAspIleProGlnPheGlyAsnThr 290
    |||||
QY 553 GTTTTCATTTGTTGTTATGGAAGGAT 579
    |||||
Db 291 CysPheSerThrGlyTyrGlyLysAsp 299

RESULT 5
P90533 PRELIMINARY: PRT: 623 AA.
AC P90533;
DT 01-MAY-1997 (Tremblrel. 03, Created)
DT 01-MAY-1997 (Tremblrel. 03, Last sequence update)
DT 01-DEC-2001 (Tremblrel. 19, Last annotation update)
DE LIMA (Fragment).
GN LIMA.
OS Dictyostelium discoideum (Slime mold).
OC Eukaryota; Mycetozoa; Dictyostelida; Dictyostelium.
OC NCBI_TaxID=44689;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=AX4;
RA Ikenfater N., Loomis W.F.;
RL Submitted (DEC-1996) to the EMBL/GenBank/DBJ databases.
CC -1- SIMILARITY: CONTAINS 3 LIM DOMAINS. THE LIM DOMAIN BINDS 2 ZINC IONS.
DR EMBL: U83086; AAB40929.1; -.
DR InterPro: IPR001781; LIM.
DR Pfam: PF00412; LIM; 2.
DR SMART: SM00132; LIM; 3.
DR PROSITE: PS00478; LIM_DOMAIN_1; 3.
DR PROSITE: PS50023; LIM_DOMAIN_2; 3.
KW LIM domain; Metal-binding; zinc.
FT NOW TRS
SQ SEQUENCE 623 AA; 71604 MW; 3DB6B8E9CF241ADE CRC64;
Alignment Scores: 1.78e-05 Length: 623
Pred. No.:

```

```

Score: 139.00 Matches: 60
Percent Similarity: 40.64% Conservative: 29
Best Local Similarity: 27.40% Mismatches: 89
Query Match: 5.21% Indels: 41
DB: 5 Gaps: 10

US-10-020-441-1 (1-1477) x P90533 (1-623)
QY 1102 CACCCCTTACACTCCACCAACAATAGTCCACATACATCAATGTCATATATTCG 1043
    |||||
Db 162 HisProGlnHis---ProHisSerHis---ProHis-GlnSerHisSerHisAl 179
    |||||
QY 1042 AGCATCA-----AGTCATCT----- 1027
    |||||
Db 179 HisSerGlnProThrThrThrValGlnProGlnGlnGlnGlnHisSerLeuSerTh 199
    |||||
QY 1026 -----CATTCACACATCTCCAAACACATCTCAACAC 992
    |||||
Db 199 ValValHisProHisAsnAlaThrProSerHisHisHisHisGlnHisSerHisSe 219
    |||||
QY 991 ATTGACCCCTACTCATGTACACAAACACATCTCCATGTTGCAATGATTTTCAT 935
    |||||
Db 219 GlnProHisSerPheAspHisAsnGlnHisGlnHisGlnProLeuGlnHisHisHis 239
    |||||
QY 934 GAGTCTCTCTCTCCAAAGTGTCTCAATCATCC--ATGACATCAATTCATCCAT 878
    |||||
Db 229 sThr-HisSerGlnProHisSerIleGlnHisGlnHisGlnValAsnLysHisIleAsp 259
    |||||
QY 877 CCAATCTACTCAATCACTCACTCACTCACTGTC--ACCTCACTCACTCACTCACTCACT 821
    |||||
Db 259 eProIleGlnHisHisValSerHisAsnIleLysAsnSerHisHisHisGlnHisGln 279
    |||||
QY 820 GAGATGAAATCACTCTCTCACTGATCACTCACTCACTCACTCACTCACTCACTCACTCACT 761
    |||||
Db 279 LHisPheSerHisSerAspLysProIleValGlnGlnHis-----V 293
    |||||
QY 760 TCAAGGCTCTGACCAATCAACAACCCGATCAGGTAGA-----CAGACAGACACAA 707
    |||||
Db 293 alGlnGlnThrThrLysThrThrThrThrGlnTyrLysProProGlnSerSerIleGln 313
    |||||
QY 706 CAGATCAACAGACACAAATTAATGATGATGATGATGATGATGATGATGATGATGATGAT 647
    |||||
Db 313 snAsn-AsnAsnAsnAsnAsnAsnLeuIleIleGlnAspSerThrHisHisSerHisAsn 332
    |||||
QY 646 TTATTCACCAACATCACTCTTTCATAT-----TTCCACATCTTCAATGATGATGAT 593
    |||||
Db 333 LeuHisAsnAsnGlnLysArgAspGlyLysSerTyrSerLysThrPheIleThrAsp 352
    |||||
QY 592 CACGGTCGTATCATCCCTTCATACCAACATGAACACACAGTTCCAG 542
    |||||
Db 353 AspProAsnTyrHisGlnAsnGlnAsnGlnHisHisSerHisGln 369

RESULT 6
Q95OR7 PRELIMINARY: PRT: 6239 AA.
AC Q95OR7;
DT 01-MAY-2000 (Tremblrel. 13, Created)
DT 01-MAY-2000 (Tremblrel. 13, Last sequence update)
DT 01-MAR-2002 (Tremblrel. 20, Last annotation update)
DE TYPE I polyketide synthase AVES 2.
GN AVEA2.
OS Streptomyces avermitilis.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
OC NCBI_TaxID=33903;
RN [1]
RP SEQUENCE FROM N.A. Pubmed=10449723;
RX MEDLINE=99380548;
RA Ikeda H., Nonomiya T., Usami M., Ohra T., Omura S.;
RT "Organization of the biosynthetic gene cluster for the polyketide
antheimintic macroide avermectin in Streptomyces avermitilis.";
RL Proc. Natl. Acad. Sci. U.S.A. 96:9509-9514(1999).
DR EMBL: AB032367; BAA84475.1; -.

```









```

Db      332 GluSerThrPro 335
      ::::|
RESULT 10
014651 PRELIMINARY; PRT; 957 AA.
AC 014651; 014650;
DT 01-JAN-1998 (Tremblrel. 05, Created)
DT 01-NOV-1998 (Tremblrel. 08, Last sequence update)
DT 01-JUN-2001 (Tremblrel. 17, Last annotation update)
DE Mucin (Intestinal mucin) (Fragment).
GN MUC3.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=97476275; PubMed=9334251;
RA Gum J.R., Jr., Ho J.J.L., Pratt W.S., Hicks J.W., Hill A.S.,
RA Vivali L.E., Robertson A.M., Swallow D.M., Kim Y.S.;
RT "MUC3 human intestinal mucin. Analysis of gene structure, the carboxyl
RT terminus, and a novel upstream repetitive region.";
RL J. Biol. Chem. 272:26678-26686 (1997).
DR EMBL; AF007194; AAC02272.1; -.
DR EMBL; AF007195; AAB84382.1; -.
DR InterPro; IPR000561; EGF-like.
DR SMART; SM00181; EGF_1.
DR PROSITE; PS00022; EGF_1; UNKNOWN_2.
DR PROSITE; PS01186; EGF_2; 1.
KW EGF-like domain; Glycoprotein.
FT NON TER
SQ SEQUENCE 957 AA; 99554 MW; E67722560D600A1P CRC64;

Alignment Scores:
Pred. No.: 9,19e-05 Length: 957
Score: 132.50 Matches: 91
Percent Similarity: 37.14% Conservative: 52
Best Local Similarity: 23.64% Mismatches: 154
Query Match: 4.97% Indels: 88
DB: Gaps: 14

US-10-020-441-1 (1-1477) x 014651 (1-957)
QY 1461 ACAAAATTCACATCTTGACCACTGACATCTGACATGATATCGGAAGTTTGAC 1402
|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 4 ThrgluThrThrSerHisSerThrProSerTyThrThrSerHisSerThr-----Thr 20
QY 1401 AGTGTGACACCATGTGATGACGACACCGAGTACTGACCTTGAAAGGATGGAGAGAGT 1342
:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 21 ThrgluThrThrProSerHisSerThrProSerTyThrThrSerHisSerThrThrThr 40
QY 1341 CCACCACTGACACCTGAGCGGTAAGTTCGAAATTCGACCTGCTTTCACACATATA 1282
|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 41 ProSerHisSerThr-----ProSerPheThr----- 49
QY 1281 GGATTGCCATTGGTGGCATGTCGCATTCATTAATAGTCGCTGACCTGTGGTCAAAC 1222
|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 50 -----SerSerHisSerThrThrThrThrThrThrThrThrHis 60
QY 1221 AAACACACACAA-----CAACAGTACAGTACAAAGACAGTACAGAT 1177
|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 61 SerThrProSerPheThrSerSerHisSerThrThrThrThrThrThrThrThrPro 80
QY 1176 GAG---AAAGAACAAACACACGAGCAAAACAAACACACACACAGTGAATTCGAAAAA 1120
|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 81 SerPheThrSerSerHisSerThrThrThrThrThrThrThrThrThrThrThrThr 100
QY 1119 TCAACTGGCACAATATCCACCTTTACCTACCCACACACACAAATAGTGCACATACATCA 1060
:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 101 Thr-----SerHisSerThrThrThrThrThrThrThrThrThrThrThrThrThr 118

```

```

QY 1059 CATGTGACATATATTCGACGATCAAGTCACTCT----- 1027
:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 119 SerHisSerThrThrThrThrThrThrThrThrThrThrThrThrThrThrThrThr 138
QY 1026 -----CCATCTCACACATCTCCAAACACATATCAACATGACCCCTACTTC 976
:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 139 ThrThrThrThrThrThrThrThrThrThrThrThrThrThrThrThrThrThrThr 158
QY 975 ATGTACACACAC-----ACTTCACCTGTTCGATAGATATTCATGAGT 931
|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 158 ThrThrThrThrThrThrThrThrThrThrThrThrThrThrThrThrThrThrThr 178
QY 930 CTCTCTCTCCCAAGTGTCTCAAT-----CATCCATCAGATCAGATTCAT 883
|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 178 ThrHisSerThrProSerSerThrThrThrThrThrThrThrThrThrThrThrThr 198
QY 882 CCATTCGATCTACTCTCACTCACTCACTCACTCACTCACTCACTCACTCACTCACTCA 823
|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 198 ProSerPheThrSerSerHisSerThrThrThrThrThrThrThrThrThrThrThr 215
QY 822 CCGAGATGGAATATCACTCTGCTACTGAACTGCAAAAGTCAATCCAAACACAGTGTAT 763
|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 215 ArgSerPheThrSerSerHisSerThrThrThrThrThrThrThrThrThrThrThr 235
QY 762 TATCAAGGCTGTGACCAATACAAACCGATCAGTAGACAGACAGAC----- 714
:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 235 ThrSerSerHisSerThrThrThrThrThrThrThrThrThrThrThrThrThrThr 255
QY 713 -----ACACAAACAGATACACACACACAGATTAATCT-----GACTACATCTGACTGA 664
|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 255 ThrThrThrThrThrThrThrThrThrThrThrThrThrThrThrThrThrThrThr 275
QY 663 CGACGAGTCAATGCTGTTATTA-----CCAACTCACTCTTCTTCATATTCACACAT 607
|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 275 ThrThrThrThrThrThrThrThrThrThrThrThrThrThrThrThrThrThrThr 294
QY 606 CTTCAGTACGAGATACCGGTGTTATCATCCCTTCATACCAACATGAACAAACACAGT 547
|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 295 ThrThrThrThrThrThrThrThrThrThrThrThrThrThrThrThrThrThrThr 314
QY 546 TCCAGTGGCGGATATCCGATGCTGTGGACACTGATCTGATTCACCTGTGAA 487
|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 314 ----- 314
QY 486 GTTGACATTTGAGCAGACATTAATGCAATATGCAATCCACTGAGTGTGCGAT 427
|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 315 -----HisSerThrProGluPheThrSerSerHisSerThrThrThrThrThr 332
QY 426 TGGTCTCTCT 418
:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 333 SerThrPro 335

```

```

RESULT 11
Q9S0R8 PRELIMINARY; PRT; 3972 AA.
ID Q9S0R8;
AC Q9S0R8;
DT 01-MAY-2000 (Tremblrel. 13, Created)
DT 01-MAY-2000 (Tremblrel. 13, Last sequence update)
DT 01-MAR-2002 (Tremblrel. 20, Last annotation update)
DE TYPE I polypeptide synthase AVE5 1.
GN AVEA1.
OS Streptomyces avermitilis.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
OX NCBI_TaxID=33903;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=9380548; PubMed=10449723;
RA Ikeda H., Nonomiyama T., Usami M., Ohra T., Omura S.;
RT "Organization of the biosynthetic gene cluster for the polypeptide
RT anthelmintic macrocyclic avermectin in Streptomyces avermitilis.";
RL Proc. Natl. Acad. Sci. U.S.A. 96:9509-9514 (1999).
EMBL; AB032367; BAA84474.1; -.

```





```

OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BERKELEY;
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brannon R.C., Rogers Y.-H.C., Blasei R.G., Champagne M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abril J.F., Aghayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA Ballew K.Y., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Botkova D., Botchan M.R., Bouck J., Brokstein P., Brotlier P.,
RA Buttis K.C., Busam D.A., Butler H., Cadien E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Dudin K.J., Evangelista C.C., Ferraz C., Ferrieria S., Fleischmann W.,
RA Foster C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibeagwu C.,
RA Jafarizadeh M., Kalush F., Karpen G.H., Ke Z., Kenson J.A., Kerchum K.A.,
RA Kimmel B.E., Kodali C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Laabo P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mettel B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milhina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusser D.R., Paclab J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spreading A.C., Stapleton M., Strong R., Sun E.,
RA Svrtkars R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yen R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of Drosophila melanogaster.";
RL Science 287:2185-2195 (2000).
DR EMBL; AF003683; AAF54394.2;
DR FlyBase; FBgn0037705; CG9381.
DR InterPro; IPR001841; Znf_ring.
DR SMART; SM00184; RING; 1.
KW Hypothetical protein.
SQ SEQUENCE 1171 AA; 126688 MW; CC6964B0B6FF2BBA CRC64;

```

```

Db 147 LysAspAlaAsnProAsnCyAspValGlnThrSerAsnSerSerglyThrCysIleTyr 166
Oy 1096 -----TTAACTATCCACAC-----AACAAATAGCCACATATCAATC 1061
Db 167 AsnAsnSerlySalahisHisHisHisAsnAsnSerAsnAsnAngly-HisThrGlnIse 186
Oy 1060 ACATGTGACATATATTCAGACAGTCAAGTCACTTCATCTCCATCCATCCATCCATCA 1001
Db 186 rHisIleHis---ThraHisSerLeuHisGlnlyHisHisHisProHisIstHrGl 205
Oy 1000 CTCMAACACATTGACCCCCCTTTCATTCATACACACACATCTCCATCGTTGATAGTA 941
Db 205 n-----MethIstHrProHisIstProHis----- 213
Oy 940 TTTCATGAGTCCTCTCTCCCTCAAGTGGTTCATATCATATCCATCATCATATTCATCC 881
Db 214 -----SerSerSerProGlnGlnSerAsnSerlyrArgGlnTyrProPr 228
Oy 880 CATCCATATCTACCTCACTCACTTCATCATCACTTCATCCATCCATCAATGACC 821
Db 228 o-----His-SerTyrSerProAsnSerProHisSerAnglyHst 242
Oy 820 GAGATGAAATCATCTCTCTCTACTGATCTCA-----CAAGTCAATCCAA 776
Db 242 hrAsnSerAsnAlaSerAsnAsnProIleSerglnArgSerAsnProGlnAlaHisProA 262
Oy 775 ACACA-----GTGTATTATCAAGCTCTGACCAATCAACAAACCG 734
Db 262 snglnAsnGlnAsnSerAsnSerAsnHetyrGluMetCySthGlyAsnSerAsnHrSerS 282
Oy 733 GATCAGGTAGACAGACACACACACACAAACAGATACACAGACACAGAAATTAAGTGA 674
Db 282 eAspGlnAsnGlnHisAsnProProProArgSerglnAsnHisHisProAsnProHis 302
Oy 673 ATCTACAGACAGACGTGATCTGCTTATTCACCAACATCACTCTTTCATATATTC 614
Db 302 IsSergln-----TyrGlnTyrGlnTyrGlnTyr 313
Oy 613 CACCATTTCTACGATCAGATCAGATCAGATGATATATCCCTTCATCAACAA 560
Db 314 ArgTyrGlnTyrHisProProHis---GlnTyrHisProLeuHisSerProGln 330
RESULT 14
Oy9552 PRELIMINARY; PRT; 3570 AA.
AC O99552;
DT 01-MAY-1997 (TREMBLrel. 03, Created)
DT 01-MAY-1997 (TREMBLrel. 03, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE Mucin MUC5B (Fragment).
GN MUC5B.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=PLACENTA;
RX MEDLINE=9716151; PubMed=9013550;
RA Desseyn J.L., Guyonnet-Duperrat V., Porchet N., Aubert J.P., Laine A.,
RA "Human mucin gene MUC5B, the 10.7 kb large central exon encodes
RT various alternate subdomains resulting in a super-repeat. Structural
RT evidence for a 11p15.5 gene family.";
RL J. Biol. Chem. 272:3168-3178 (1997).
DR EMBL; Z72496; CAA96577.1; -.
FT NON_TER 1
FT TER 3570
SQ SEQUENCE 3570 AA; 361019 MW; DE04EAD727579312 CRC64;

```

```

Alignment Scores:
Pred. No.: 0.000796 Length: 3570
Score: 124.50 Matches: 105
Percent Similarity: 31.54% Conservative: 36

```

Best Local Similarity: 23.49%  
 Query Match: 4.67%  
 DB: 4  
 Mismatches: 146  
 Indels: 160  
 Gaps: 22

US-10-020-441-1 (1-1477) x Q99552 (1-3570)

QY 1455 TCACAACTTTGACCACTGGCATATCGCAAGATATCGGAGAGTTGGCAGTGTG 1396  
 DB 695 SerThrValLeuThrThrThrAlaThrThrThr-----GlyAlaThrGlySerVal 711  
 QY 1395 ---ACACCATGTGATACG----- 1381  
 DB 712 AlaThrProSerSerThrProGlyThrAlaHisThrThrLysValLeuThrThrThr 731  
 QY 1380 -----ACACCGAGTACTGGACCTTGAGGATGGAGAGAGTCCA--- 1339  
 DB 732 ThrGlyPheThrAlaThrProSerSerSerProGlyArg---AlaArgThrLeuProVal 750  
 QY 1338 -----CCACTGTCACCTGGAGCGGGTAACTGTCCGAATTTCTGA 1300  
 DB 751 TrpLeSerThrThrThrThrProThrThrArgGlySerThrValThrProSerSerile 770  
 QY 1299 CCTGCTTCACACATATGAGATTGGCATGTGCGATGTCGCAATTCATTATGTCGT 1240  
 DB 771 ProGlyThrThrHis-----ThrProThrValLeuThrThrThrThrThrValAla 788  
 QY 1239 CGACCTGTGTGCAACCAACCAACCAACAGTACAGT----- 1198  
 DB 789 ThrGlySerMetAlaThrProSerSerSerThrGlnThrSerGlyThrProProSerLeu 808  
 QY 1197 ACAACAACTGACATGACAGATGAGAAAGAAACAAACACACGGAACAAACAAACA--- 1141  
 DB 809 ThrThrThrAlaThrThrThrThrThrAlaThrThrGlySerThrThrAsnProSerSerThrPro 828  
 QY 1140 ---CACAAAGTAATTTCAAAATATCACTGGCACAATATCCACC----- 1099  
 DB 829 GlyThrThrProLeuProLeuProValLeuThrThrThrAlaThrProAlaAlaThrSer 848  
 QY 1098 -----CTTACACTACCCACACACAAACAAATAGTCCACATACAA 1063  
 DB 849 SerThrValThrProSerSerSerAlaLeuGlyThrThrHisThrThrProProVal----- 865  
 QY 1062 TCACATGTGCACTATATTCGACGATCAAGTCACTCCATCTCACCACCTGCCAACACA 1003  
 DB 866 -----ProAsn-ThrThrAlaThrThrHis 873  
 QY 1002 TACTCAACA-----CATTAACCCCTTACTTCATATCAACACACACTTCATCGT 952  
 DB 873 sGlyArgSerLeuSerProSerSerProHisThrValCysThrAlaThrThrSerAlaThr 893  
 QY 951 T-----TGCATAGATATTTCAATGATCTCTCTCCAAAGTGGTCAATCATCCC-- 900  
 DB 893 rSerGlyLysLeuGlyThrThrHisHisThrGlnProSerThrThrGlyThrSerHisThrPr 913  
 QY 899 -----ATCACATCAATTCATCCCATCAATCAATCAATCAATCAATCAATCAATCA 856  
 DB 913 AlaAlaThrThrThrGlyThrThrGlnHisSerThrProAla----- 926  
 QY 855 ACATACGCTCACTCACTCACTCACTCACTCACTCACTCACTCACTCACTCACTCACTCA 796  
 DB 927 -----LeuSerSerProHisPro-----SerSerArgThrThr 937  
 QY 795 ACTGTCACAAAGTATCCAAACACAGTTGTATATCAAGGCTGACCAATACAAACAC 736  
 DB 937 rGluSerProProSerProGlyThrThrThr-----ProGlyHisThrThr 952  
 QY 735 CGGATCAGTATGACACACACACAAACAGATACACACACACAGATATGACTG 676  
 DB 952 rAlaThrSerArgThrThrAlaThrAlaThrAlaThrProSerLysThrArgThrSerThr----- 970  
 QY 675 ACATCTGACTGACGACTGATGATGCTTATATCAACAACTCACTTCTTCAATAT 616  
 DB 971 -----LeuLeuProSerGlnProThr----- 977

QY 615 TCACACATTTTACGTGACGATACGGTCTGTATCATCCCTCCATACCAACAAATGAA 556  
 DB 978 -----SerAlaProLeuThrThrValVa 985  
 QY 555 AACA----- 552  
 DB 985 lThrMetGlyCysGluProGlnCysAlaThrPserGlnTrpLeuAspTyrSerTyrProMe 1005  
 QY 551 -CCAGTTCAGGTGGCGGAGATATCCGAT-----GGCTGTGGCAGACT 511  
 DB 1005 cProGlyProSerGlyGlyAspPheAspThrTyrSerMetIleArgAlaIaGlyGlyAl 1025  
 QY 510 GATCACTCTGATTCACATC-----TGTAACTTGACATTTGAGCCACATTAACAT 460  
 DB 1025 aValGlyGluGlnProLeuGlyLeuGluCysArgAlaGlnIaGlnProGlyValProLe 1045  
 QY 459 TGCATATATGCAATCCACTGAGTGTCTGTCGATT-----GGTCTCTCTGTCTGCGCA 406  
 DB 1045 uArgGluLeuGlyGlnValValaGluCysSerLeuAspPheGlyLeuValCysArgAsnAr 1065  
 QY 405 CATACAGAGGCGCATGTAT 387  
 DB 1065 gGluGlnValGlyLysPhe 1071  
 RESULT 15  
 Q29464 PRELIMINARY; PRT; 237 AA.  
 AC Q29464.  
 DT 01-NOV-1996 (TREMBlurel. 01, Created)  
 DT 01-NOV-1996 (TREMBlurel. 01, last sequence update)  
 DT 01-JUN-2002 (TREMBlurel. 21, last annotation update)  
 DE Trypsase (EC 3.4.21.59) (Fragment).  
 OS Bos taurus (Bovine).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;  
 OC Bovidae; Bovinae; Bos.  
 OX NCBI\_TaxID=9913;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=LIVER CAPSULE;  
 RX MEDLINE=96203914; PubMed=8620861;  
 RA Pallao M., Gambacurta A., Fiorucci L., Mignogna G., Barra D.,  
 RA Ascoldi F.;  
 RT "cDNA cloning and primary structure of trypsin from bovine mast cells  
 and evidence of the expression of bovine pancreatic trypsin inhibitor  
 mRNA in the same cells.";  
 RT Eur. J. Biochem. 237:100-105(1996).  
 CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE  
 CC TRYPSIN FAMILY.  
 DR EMBL; X94982; CA64438.1; -.  
 DR HSSP; P20231; LMO.  
 DR MEROPS; S01.118; -.  
 DR InterPro; IPR001314; Chymotrypsin.  
 DR InterPro; IPR001254; Ser protease\_Try.  
 DR Pfam; PF00089; trypsin.1.  
 DR PRINTS; PR00722; CHYMOTRYPSIN.  
 DR SMART; SM00020; TRYP\_SBC; 1.  
 DR PROSITE; PS50240; TRYPSIN\_DOM; 1.  
 DR PROSITE; PS00134; TRYPSIN\_HIS; UNKNOWN\_1.  
 KW Hydrolyase; Serine protease.  
 FT NON TER  
 FT 1  
 SQ SEQUENCE 237 AA; 26550 MW; EA537A75294EFBBA CRC64;  
 Alignment Scores:  
 Pred. No.: 0.000556 Length: 237  
 Score: 124.00 Matches: 41  
 Percent Similarity: 38.15% Conservative: 25  
 Best Local Similarity: 23.70% Mismatches: 43  
 Query Match: 4.62% Indels: 64  
 DB: 6 Gaps: 5  
 US-10-020-441-1 (1-1477) x Q29464 (1-237)

Oy	82	AGACAACTGTATACAGTTTCCTAGTCTCAACAGACACTATCACTAGCTGCTATTGG	141
Db	18	ArgHisHisCysGlyGlySerLeuIleHisProGlnTrpValLeuIleIleIleIleHisCys	37
Oy	142	GTTTGCCTCACCATTG-----CCAGTGAATTGGGGTAAAGAGATGACATGAACACATT	192
Db	38	ValGlyProGluValHisGlyProSerTyrPheArgValGlnLeuArgGlnHisIleu	57
Oy	193	GTTGCTCAATGTAATTCGATTTGATTTCACAGACAGTGTGGTGCGTGCGCTTTGGTT	252
Db	58	TyrTrpGlnAspGln-----	62
Oy	253	TGTGTATCAGTGGGTTTGCATTAAGATTTTAATGTCTATATATTGTTTCAAGTT	312
Db	63	-----LeuLeuProIle	66
Oy	313	TCATTTTCACACTGAGGAATGGCGACCAACAAGCATCCATCCAAACCGTCTGGAGTT	372
Db	67	Ser-----	67
Oy	373	AAGTGGCACCAGATACATGCGCTCTGTATGTGGCAGCAGACAGAGACCAATGC	432
Db	68	ArgIleIlePro-----HisProAsnCysIleSerValIys-----	79
Oy	433	CAGACACTCAGTGGATTCGATTTGCATTTGTAAATGCTGGCTCAATGTCAACTTACAG	492
Db	80	-----AsnGlyAlaAspIleIleIleuSerGlnIleuAspIleuValLeuIleSer	96
Oy	493	AGTGAATCAGAGTATCAGTCTGCGACACAGCATCGATATCCGCCAAGCTGGAACTGGT	552
Db	97	TrpHisValGlnProValThrLeuProProGluSerGlnThrPheProProGlyThrGln	116
Oy	553	GTTTTCATGTTGGTTATGGAAGGATGATACACCGT	591
Db	117	CysTrpValThrGlyTrpGlyAsnValAspAsnGlyIle	129

RESULT 16			
ID	018459	PRELIMINARY;	PRT; 272 AA.
AC	018459;		
DT	01-JAN-1998	(TREMBLrel. 05, Created)	
DT	01-JAN-1998	(TREMBLrel. 05, Last sequence update)	
DT	01-JUN-2002	(TREMBLrel. 21, Last annotation update)	
DE	Serine proteinase precursor.		
GN	SP-III.		
OS	Heterodera glycines (Soybean cyst nematode).		
OC	Eukaryota; Metazoa; Nematoda; Chromodorea; Tylenchida; Tylenchina;		
OC	Tylenchoidea; Heteroderidae; Heteroderinae; Heterodera.		
OX	NCBI_Taxid=51029;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RX	MEDLINE=968030247;	PubMed=9364965;	
RA	Lilley C.J., Utwin P.E., Atkinson H.J., McPherson M.J.;		
RT	"Characterization of cDNAs encoding serine proteinases from the		
RT	Soybean cyst nematode Heterodera glycines.";		
RL	Mol. Biochem. Parasitol. 89:195-207(1997).		
CC	-!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE		
CC	TRYPSIN FAMILY.		
CC	EMBL; Y13906; CAA74204.1; -.		
DR	HSSP; P20231; IMAO.		
DR	InterPro; IPR001314; Chymotrypsin.		
DR	InterPro; IPR001254; Ser_protease_Try.		
DR	Pfam; PF000089; trypsin; 1		
DR	PRINTS; PR000722; CHYMOTRYPSIN.		
DR	SMART; SM00020; Tryp_Spc; 1.		
DR	PROSITE; PS50240; TRYPSIN_DOM; 1.		
DR	PROSITE; PS00134; TRYPSIN_HIS; UNKNOWN_1.		
DR	PROSITE; PS00135; TRYPSIN_SER; 1.		
KW	Hydrolase; Serine protease; Signal.		
FT	SIGNAL	1	22
FT	CHAIN	23	272
FT	SEQUENCE	272 AA;	30194 MW; 1P22SC2635DF1P9 CRC64;

Alignment Scores:	
Pred. No.:	0.000642
Score:	123.50
Percent Similarity:	35.86%
Best Local Similarity:	21.91%
Query Match:	4.60%
DB:	5
Length:	272
Matches:	55
Conservative:	35
Mismatches:	84
Indels:	77
Gaps:	10

US-10-020-441-1 (1-1477) X 018459 (1-272)

[illegible]







QY 1377 ---CCGAGTACTGACCTTGAAGGATGGAGAGAGTCCACCACTG-----TCA 1330  
 Db 108 ValAlaHisThrAsnSerAlaThrSerSerArgProProProPheThrThrHisSer 127  
 QY 1339 CCTGAGAGGGGATGACTGTCGAAATTTGACCTGCTTACACATATAGATTTGCCATTG 1270  
 Db 128 ProProThrGlySerSerPro-----PheSerSerThrGly---Promet 141  
 QY 1269 GTCCGATGCGGCATTTCATTATAGTCGCTGACCCCTGTGGTGAACAAACACACACA 1210  
 Db 142 ThrAlaThrSerPheLys----- 147  
 QY 1209 CAACAGTACAGTACAAACAGTACAGTACAGTACAGTACAGTACAGTACAGTACAG 1150  
 Db 148 -----ThrThrThrThrThrThrThrThrThrThrThrThrThrThrThrThr 162  
 QY 1149 AACAAACACACACAGTATTTCCAAATATGACTGGACAAATATCCACCTTTACACT 1090  
 Db 163 ---LeuThrHis-----ValProProPheSerThrSerLeuValThrProle----- 177  
 QY 1089 ACCGACACACAAATA-----GTCCACATACATACATACATACATACATATTTGAGAGATCA 1036  
 Db 178 ThrHisThrValIleThrProThrHisProGlnMetSerThrSerAlaGlyIleHisSer 197  
 QY 1035 AGTCACTCTCCATCTCACCACCTCTCCAAACACATACATACATACATACATACATCTTC 976  
 Db 198 ThrProThrGlyThrIleAlaSerProThrThrValIleValThrAlaGlySerThrGlyThr 217  
 QY 975 -----ANGTACACACACACATCTCCATGCTTTGATGATATTTATGAGTCTTCTCC 922  
 Db 218 AlaProLeuMetThrAlaThrThrSerArgIleSerGlnAlaHisSerSerIleSerThr 237  
 QY 921 TCCACAGTGTGCATCATCCCAT-----CACATACATCTCCATCC--ATCCAA 874  
 Db 238 AlaLysThrSerThrSerLeuHisSerHisAlaSerSerThrHisHisProGlnValThr 257  
 QY 873 TCTACCTCACT-----CACTCACATCAGCT----- 847  
 Db 258 ProThrSerThrThrAsnValThrProLysSerThrSerArgAspThrSerThrProVal 277  
 QY 846 ---CACTCACTCTCACCACACATGACCGAGAT-----GGAACTACTCT-- 804  
 Db 278 ThrHisThrThrSerAlaThr--SerSerArgProProThrProIleThrThrHisSerSe 297  
 QY 803 -----GCTAGTGAAGTGTCAAAAGTATCCAAAGTGTATATATCAAGGC 754  
 Db 297 rProThrArgSerSerProLeuSerSerThrGlyProMetThrAlaThrSerIleLysThr 317  
 QY 753 TCTGACCAATACAAACACCGGATCAGTACAGACAGACACACAAACAGATACA----- 699  
 Db 317 rThrThrThrThrThrProThrProSerHisProGlnThrThrLeuThrThrHisValProBr 337  
 QY 698 ---CAGACACACAGATTAAGTACTGACATGACATGACATGACATGACATGACATGAC 643  
 Db 337 oPheSerThrSerSerValThrProSerThrHisThr-----ValIleThrProth 354  
 QY 642 TCACCAACACATCTTCTTCAATATCCACCATTTCTTACGTGACGAGTACCGTCTGT 583  
 Db 354 rHisIleGlnMetSer-----ThrSerAlaSerIle 364  
 QY 582 ATCATCCCTTCCA-----TAACCAACATGAAACACACAGTCTCCAGTGGCGGAT 532  
 Db 364 eHisSerThrProThrGlyThrValProProLeuThrThrArgMetProThrGlySerThr 384  
 QY 531 ATCCGATGCTGTGGCAGATGATGATCTGATTCACCTCTGTAAG 486  
 Db 384 rArgThrGlyProProMetThrGlyThrIleIleGlnThrSerLys 399  
 RESULT 21  
 076898 PRELIMINARY; PRT; 281 AA.  
 AC 076898;  
 DT 01-NOV-1998 (TREMBlrel. 08, Created)

DT 01-NOV-1999 (TREMBlrel. 12, last sequence update)  
 DE 01-MAR-2002 (TREMBlrel. 20, last annotation update)  
 DT EG:80H7.1 protein.  
 GN EG:80H7.1  
 OS Drosophila melanogaster (fruit fly).  
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
 OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
 OC Ephydroidea; Drosophilidae; Drosophila.  
 OX NCBI\_TaxID=7227;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Mottler S., Cadieu E., Dreano S., Leilaure V., Galibert F.;  
 RT "Sequencing the distal X chromosome of Drosophila melanogaster."  
 RL Submitted (JUL-1998) to the EMBL/GenBank/DBJ databases.  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RA Benoit P.;  
 RL Submitted (APR-1999) to the EMBL/GenBank/DBJ databases.  
 CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE  
 CC TRYPSIN FAMILY.  
 DR EMBL; AL031027; CA119843.2; -.  
 DR HSSP; P00766; 1CHG  
 DR FlyBase; FBgn025385; EG:80H7.1.  
 DR InterPro; IPR001314; Chymotrypsin.  
 DR InterPro; IPR001254; Ser\_protease\_Try.  
 DR Pfam; PF00089; trypsin.1.  
 DR PRINTS; PR00722; CHYMOTRYPSIN.  
 DR SMART; SM00020; TRYPSIN\_DOM.1.  
 DR PROSITE; PS0240; TRYPSIN\_DOM.1.  
 DR PROSITE; PS00134; TRYPSIN\_HIS; UNKNOWN\_1.  
 DR PROSITE; PS00135; TRYPSIN\_SER; 1.  
 KW Hydroxylase; Serine protease.  
 SQ SEQUENCE 281 AA; 29875 MW; 4E78B0A6031949B CRC64;  
 Alignment Scores:  
 Pred. No.: 0.00118 Length: 281  
 Score: 121.00 Matches: 51  
 Percent Similarity: 34.55% Conservative: 25  
 Best Local Similarity: 23.18% Mismatches: 70  
 Query Match: 4.50% Indels: 74  
 DB: Gaps: 7  
 US-10-020-441-1 (1-1477) x 076898 (1-281)  
 QY 49 GAATCCCGGTCATCGCATTTCTTA-----ACGACAGAGACAATGTCACAGST 99  
 Db 37 GluPheProPheValValSerLeuArgAlaLysSerGlyArgHisSerCysGlyAla 56  
 QY 100 TCACTAGTCTCAACGAGACGACTCTCACAGTGTGCTTGTGCTCACCATTGCCA 159  
 Db 57 ThrIleuAsnProCysTrrValIleuThrAlaHisCysValArgGlySerSerPro 76  
 QY 160 GTGATTCGGGTAAAGATGACGACGAAACACATGTTGCTCAATGTAATGATTTC 219  
 Db 77 Glu----- 77  
 QY 220 CAGACAGTGTTCGCTGCTGCTGCTTGTGTGTATGACAGTGGTTGCATAGATT 279  
 Db 77 ----- 77  
 QY 280 TTAATGTATATACCTTGGTCTTATTTACAGTTTCATTTCTCACACTGAGGATGGAGC 339  
 Db 78 -----GlnIleuAspIleuGlnTrrGlySerGlnMetLeuAlaArgAsnSerSer 93  
 QY 340 CAACAGGACATTCATCCACAAACGCTGTGAGTAAAGTGGACGACAGATTAACATCCCTCT 399  
 Db 94 GlnValAla-----ArgValAlaIleAlaIlePheValHisProGlyTrrGlyProGlu 110  
 QY 400 TGTATGCGCAGACAGACAGAGACCAATGCCACAGACATCATGATGATTCGATTTGCA 459  
 Db 111 AspLysTrrValAsn-----AspIleAla 118  
 QY 460 ATTGAATGCTGGCTCAAATGTCATCTTACAGATGGAATCAGAGTATCAGTCTGCA 519

```

Db 119 LeuLeuGlnLeuAlaGlnSerValAlaLeuSerLysPheValGlnProValArgLeuPro 138
      ::::: ||||| ||| ::::: ::::: |||||
Oy 520 CAGCCATCGAATGCCCGCCACCTGGAACTGGTGTTCATGTTGATGAGGAGGAT 579
      ::::: ||||| ::::: ||||| ::::: |||||
Db 139 GlnProArgGlnValThrProGlnAsnAlaSerAlaValLeuAlaGlyTyrPoliLeuAsn 158
      ::::: ||||| ::::: ||||| ::::: |||||
Oy 580 GATTAACGACCGTATCCGTACGTAAAGATGGTGGATA-----TTGAAGAA 627
      ::::: ||||| ::::: ||||| ::::: |||||
Db 159 -----AlaThrGlyGlyValValGlnGlnHisLeuGlnLys 170
      ::::: ||||| ::::: ||||| ::::: |||||
Oy 628 GTGAGTGTGGTGAATTAACGACATGACTGCTGCTGATGATGATGACAGTTAT 687
      ::::: ||||| ::::: ||||| ::::: |||||
Db 171 ValLysLeuGlnValPheSerAsp---ThrGlnCysSerGlnArgHisGlnThrTyrLeu 189
      ::::: ||||| ::::: ||||| ::::: |||||

RESULT 22
Oy 097366 PRELIMINARY; PRT; 365 AA.
AC 097366;
DT 01-MAY-1999 (TRENBLrel. 10, Created)
DT 01-MAY-1999 (TRENBLrel. 10, Last sequence update)
DE 01-MAR-2002 (TRENBLrel. 20, Last annotation update)
DE Pro-phenoloxidase activating enzyme-I precursor.
OS Holotrichia diomphalia.
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Coleoptera; Polyphaga;
OC Scarabaeiforma; Scarabaeidae; Melolonthinae; Holotrichia.
OX NCBI_TaxID=33394;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=99053686; Pubmed=9839951;
RA Lee S.Y., Cho M.Y., Hyun J.H., Lee K.M., Homma K., Natori S.,
RA Kawabata S., Iwanaga S., Lee B.L.;
RT "Molecular cloning of cDNA for pro-phenol oxidase-activating factor I,
RT a serine protease that induced by lipopolysaccharide or 1,3-beta-
RT glucan in coleopteran insect, Holotrichia diomphalia larvae.";
RL Eur. J. Biochem. 257:615-621(1998).
CC -! SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE
CC TRYPSIN FAMILY.
DR EMBL: AB013088; BAA34642.1; -.
DR HSRP: P00763; IDPO.
DR MEROPS: S01.204; -.
DR InterPro: IPR001314; Chymotrypsin.
DR InterPro: IPR001254; Ser protease_Try.
DR Pfam: PF00089; trypsin.1.
DR PRINTS: PR00722; CHYMOTRYPSIN.
DR SMART: SM00020; TRY_P_SPC.1.
DR PROSITE: PS50240; TRYPsin_DOM; 1.
DR PROSITE: PS00134; TRYPsin_HIS; UNKNOWN_1.
DR PROSITE: PS00135; TRYPsin_SER; 1.
KW Hydrolyase; Serine protease; Signal.
FT SIGNAL 1 21 POTENTIAL.
FT SIGNAL 22 109 POTENTIAL.
FT CHAIN 110 365 POTENTIAL.
SQ SEQUENCE 365 AA; 40194 MW; EC69F6093923F4C5 CRC64;

Alignment Scores:
Pred. No.: 0.00124 Length: 365
Score: 121.00 Matches: 44
Percent Similarity: 39.46% Conservative: 29
Best Local Similarity: 23.78% Mismatches: 64
Query Match: 4.50% Indels: 48
DB: 5 Gaps: 7

US-10-020-441-1 (1-1477) x 097366 (1-365)
Oy 49 GAATTCCTTCATCGCATTTCTTA-----ACGACAGAGACAACATG 90
      ||||| ||||| ::::: |||||
Db 120 GlnPheProThrThrAlaMetIleGlyTyrLysAsnSerSerAsnPheGlnPheAla 139
      ||||| ||||| ::::: |||||
Oy 91 TGTACAGSTTCATAGTCTCAACAGAGACGATCTACAGCTGTGATGTGTGGTCTCA 150
      ||||| ||||| ::::: |||||
Db 140 CysGlyGlySerLeuIleAsnArgTyrIleValThrAlaHisCysValaIagly 159
      ||||| ||||| ::::: |||||

```

```

Oy 151 CCATTGCCAGTATTCGGGTGAAGAGATGACTGAAACACTT---GTTCTCATGTAT 207
      ||||| ||||| ::::: |||||
Db 160 -----ArgValIleuArgValValGlyAlaLeuAsnLysValArgLeuGlyGlnTrpAsn 177
      ::::: ||||| ::::: |||||
Oy 208 CGATTGATTTTCAACAGACAGCTGTTTCCGTGCGCGTTCGTTGTGTATGACGTGG 267
      ::::: ||||| ::::: |||||
Db 178 ThrAlaThrAspProAspCysTyrGlyAlaValArgValCys----- 191
      ::::: ||||| ::::: |||||
Oy 268 TTTGCATGATTTTATATGCTATATCTTGGCTTATTTACGGTTTCATTTCTCACACTG 327
      ||||| ||||| ::::: |||||
Db 192 -----ValProAspLysProIle 197
      ::::: ||||| ::::: |||||
Oy 328 AGGAATGGCGAACCAAGGATTCATCCACCAACCGTGTGAGTTAAGTGGCACCGAGA 387
      ||||| ||||| ::::: |||||
Db 198 AspLeuGlyIleGlnGlnThrIleGlnHis-----ProAsp 209
      ::::: ||||| ::::: |||||
Oy 388 TACATGCCCTCTTGTATGTCCGACAGACAGAGAGAACCAATCCGACACAGACTCAGTGA 447
      ||||| ||||| ::::: |||||
Db 210 TyrValAsp-----GlySerLysAspArgTyr----- 218
      ::::: ||||| ::::: |||||
Oy 448 TTCGATTTGCAATTTGATGCTGCTCAAAATGTCACATTACAGATGGAATCAGAGTG 507
      ||||| ||||| ::::: |||||
Db 219 HisAspIleAlaLeuIleArgLeuAsnArgGlnValGlnPheThrAsnTyrIleArgPro 238
      ::::: ||||| ::::: |||||
Oy 508 ATCAGCTTGCACACAGCATCGGATATCCCGCACCTGGAAGTGTGTTTCATTGTTGGT 567
      ::::: ||||| ::::: |||||
Db 239 ValCysLeuProGlnProAsnGlnGlnValGlnValGlyGlnArgLeuThrValValGly 258
      ::::: ||||| ::::: |||||
Oy 568 TATGAGAGGAGATGAT 582
      ::||| ||||| ::::: |||||
Db 259 TrpGlyArgThrGlu 263
      ::||| ||||| ::::: |||||

RESULT 23
Oy 014887 PRELIMINARY; PRT; 477 AA.
ID 014887;
AC 014887;
DT 01-NOV-1996 (TRENBLrel. 01, Created)
DT 01-NOV-1996 (TRENBLrel. 01, Last sequence update)
DE 01-DEC-2001 (TRENBLrel. 19, Last annotation update)
DE Mucin (Fragment).
GN MUC5AC.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=TRACHEOBRONCHIAL MUCOSA;
RA Guyonnet-Duperrat V., Audie J., Debailleu V., Laine A., Buissine M.,
RA Zoultina-Gallie S., Pigny P., Aubert J., Porchet N.;
RT "Characterization of the human mucin gene MUC5AC: a consensus
RT cyteine-rich domain for 11p15 mucin genes.";
RL Biochem. J. 0:0-0(1994).
DR EMBL: Z34277; CAA84031.1; -.
FT NON_TER 1 1
FT NON_TER 477 477
SQ SEQUENCE 477 AA; 50666 MW; E7360031B4A51311 CRC64;

Alignment Scores:
Pred. No.: 0.0013 Length: 477
Score: 121.00 Matches: 104
Percent Similarity: 34.59% Conservative: 34
Best Local Similarity: 26.07% Mismatches: 150
Query Match: 4.54% Indels: 111
DB: 4 Gaps: 14

US-10-020-441-1 (1-1477) x 014887 (1-477)
Oy 1444 AGCCACACATCGATCTCGACATATATCGGGAAGTTTGGCAGTGTGAC----- 1394
      ||||| ||||| ::::: |||||
Db 75 SerHisProGlnValSerIleGlnHisLeuGlyGlnValValGlnCysSerArgGlnGlu 94
      ||||| ||||| ::::: |||||

```

```

Qy 1393 -----ACATGTGATACGAC 1379
Db 95 GlyLeuValCySArgaEnglnaBpGlnGlnGlyProPhelyMetCysLeuasnTyrGlu 114
Qy 1378 ACCGAGTACTGG-----ACCTTGAAGGATGGAGAGAGGTCCACCACTGTCC 1328
Db 115 ValArgValLeuCySArgGlnUthrProlyGlyCysProValThrSerThrProValThr 134
Qy 1327 TGGAGCGGGTAACCTGCCGAATTCCTGACCTCTTACACATATAGATTCGATGGT 1268
Db 135 AlaProSerThrProSer-----Gly 141
Qy 1267 CGCATGTGGCATTCATTAATGATGCTGCAGCCCTGGTGCAACAAACACACAA 1208
Db 142 ArgAlaThrSerProThrGlnSerThrSerSer---TpgGlnUthrSerArgThrThr 160
Qy 1207 ACAGTACAGTACACAGTACAGTACAGTACAGTACAGTACAGTACAGTACAGT 1148
Db 161 LeuVal-ThrThrSerThrThrSerThr-----ProGlnThrSerThrThrSerAl 177
Qy 1147 CAAACACACAAAGTAATTTGCAAAATATGCACTGGCACAATATCCACCTTACATAC 1088
Db 177 AProThrThrSerThrThrThrSerAlaProThrThrSerThrThrSerAla---ProThrTh 196
Qy 1087 CCACACAAACAATAGTCCACATACATACATGACATATATTCGACGATCAAGTCACTC 1028
Db 196 rSerThrThr-----SerThrProG 203
Qy 1027 TCATCTCACCACCTTCCAAACACATCTGCAACACATTTGACCCCTTACTTATGATCA- 969
Db 203 nThrSerThrSerSerAlaProThrSerSerThrThrSerAla-ProThrSerSerThrTh 223
Qy 968 -----CAACACACTTCACCTGCTTGACATAGATATTTGATGAGTCTCTCTCCAA 917
Db 223 leSerAlaArgThrThrSerThrThr----- 231
Qy 916 GTGGTTCATATCCATCCATCAATCATTCATCCATCCATCAATCTACCTCACTCACT 857
Db 232 -----SerAlaProThrThrSer-----ThrThrs 240
Qy 856 CACATCAGTCACTCACTCACTCACTCACTCACTCACTCACTCACTCACTCACTG 797
Db 240 erPheProThrThrSerThrThrSerAlaThrThrThrThrThrSerThrSerAlaProThrs 260
Qy 796 AACTGTCAAAAGTCACTCACTCACTCACTCACTCACTCACTCACTCACTCACTCA 737
Db 260 erSerThrThrSerThrProGlnThrSerLysThrSerAlaAlaThrSerSerThrThrs 280
Qy 736 CCGATCAGTACAGACAGACAGACAGACAGACAGATACACAGACAGATACAGTCACT 677
Db 280 erGlySerGlyThrThrProSerProValThrThrThrThrThrThrThrThrThrThr 299
Qy 676 GACATCTGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACGAC 617
Db 300 -ThrSer-----ThrSerThrSerThrSerThrSerThrThrThrThrThrThrThr 316
Qy 616 TTCACCACTTCTTACGTACGAGTACG-----GTCCGTATATATCCCTTC 572
Db 316 lThrArgAspCysHisProArgCysThrThrThrThrThrThrThrThrThrThrThrThr 336
Qy 571 C-----ATACCAACCAATGAA 554
Db 336 rProGlyProHisGlyGlyAspGlyUthrThrThrThrThrThrThrThrThrThrThrThrThr 356
Qy 553 CACCACTTCAGTGGCGGATATCCAGTGGTGGCAGACTGATCACTGTGATTCAC 494
Db 356 sileCysArgArgProGlnGlnUthrThrArgGlnUthrCysArgAlaLysSerHisProG 376
Qy 493 TCTGTAATGACATTTGAGCAGCATTAACAATGCAATATGACATGACATGACATGAC 434
Db 376 uValSerThrThrThrThrThrThrThrThrThrThrThrThrThrThrThrThrThrThr 391
Qy 433 GTGGATTTGCTCTCTGTCTGTCTGTCCGACATACAGAGGCGATGAT 387

```

```

Db 392 ----Glu-GlyLeuValCySArgaEnglnaBpGlnGlnGlyProPhe 405
RESULT 24
ID 09UKNO PRELIMINARY; PRT; 957 AA.
AC 09UKNO;
DT 01-MAY-2000 (TREMBlrel. 13, Created)
DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)
DT 01-MAY-2002 (TREMBlrel. 20, Last annotation update)
DE Mucin 11 (Fragment).
GN MUC11.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=COLON MUCOSA;
RX MEDLINE=99391252; PubMed=10463611;
RA Williams S.J., McGuckin M.A., Gocley D.C., Eyre H.J., Sutherland G.R.,
RA Antalis T.M.;
RT "Two novel mucin genes down-regulated in colorectal cancer identified
RT by differential display.";
RL Cancer Res. 59:4083-4089(1999).
DR EMBL; AF147791; AAD5679.1; -.
DR InterPro; IPR001589; Actbind_actnin.
DR PROSITE; PS00019; ACTININ_1; UNKNOWN_1.
FT NON TER
FT NON TER
FT NON TER
SQ SEQUENCE 957 AA; 96261 MW; 3127BFA3D3A77C11 CRC64;

Alignment Scores:
Pred. No.: 0.00146 Length: 957
Score: 121.00 Matches: 118
Percent Similarity: 30.29% Conservative: 38
Best Local Similarity: 22.91% Mismatches: 179
Query Match: 4.54% Indels: 180
DB: Gaps: 24

US-10-020-441-1 (1-1477) x 09UKNO (1-957)
Qy 1366 ACCTTGAAGGATGGAGAGAGGTCCACCACTGCA-----CCTGAGCGGGTAA 1316
Db 344 ThrSerSerGlyValSerGlnUthrThrThrThrThrThrThrThrThrThrThrThrThr 363
Qy 1315 CTGTCCGAATTCGACCTGCTTTCACACATATAGATTCGATTCGATTCGATTCGCA 1256
Db 363 sThrThrAlaPheProAspSerThrThrThrProGlyLeu-----SerArgH 379
Qy 1255 TTCATT-----ATACTCGCTGACCGCTGCTGC 1226
Db 379 sSerThrThrSerHisSerSerProGlySerThrAspThrThrLeuLeuProAlaSerTh 399
Qy 1225 AAACAAACACACACACACAGTACAGTACAAACAGTACAGTACAGTACAGTACAGT 1166
Db 399 rThrThrSerGlyProSerGlnUthrThrThrThrThrThrThrThrThrThrThrThrThr 419
Qy 1165 AAACAAACGGAACAAACAAACAAACACACAGTAAATTCGAAATATGACACTGCACAT 1106
Db 419 pThrAlaLeuSerProGlySerThr-----ThrAlaLeuSerPheGlyGlnG 435
Qy 1105 ATCACCCCTTAC-----ACTACCCACACACA----- 1078
Db 435 uSerThrThrThrThrThrThrThrThrThrThrThrThrThrThrThrThrThrThrThr 455
Qy 1077 -----ATAGTCCACATACATACATGCGCATATATTCGACGATCAAGTCA 1031
Db 455 rThrSerSerGlyLeuAlaGlnUthrThrThrThrThrThrThrThrThrThrThrThrThr 472
Qy 1030 CTCTCCATTCACCACTGCTC-----C 1010
Db 472 ySerProArg-ThrThrThrThrThrThrThrThrThrThrThrThrThrThrThrThrThrThr 492

```

OY	1009	AAACCAATACGTAACAACATTGGACCCCGCTTCATCATGTAGACAACAACACTTCCATGCATT	95
Db	492	eethAlaIphelInThrHisProAlaSerThrHisTrpThrProSerThr-----	508
OY	949	GCAATAGATATTTTCATGAAGTCTTCTCCGCCAAGTGCTCATAATCATCCCATC-----	897
Db	509	-----ProSerThrAlaThrAlaProValIGluLus	519
OY	896	----ACATCATTTCCATCCCATCCCATCTCACTCAACTCACTCATC-----	852
Db	519	eethTrhThyHisArgSerProSerSerThrPro-----ThrTrhHisPheProAlas	537
OY	851	--CACGTCACTCCATCCACCACA-----CACAAATGACCGGATGAGAA	812
Db	537	erSerThrThrseryGlyHisserGlyLysSerThrIlePheHisSerSerProAspAla-	556
OY	811	ATCACTGCTGTAAGAATGTGTCACAAAAGTCATCCAACAACAGTTGTTATTATCAAGGCTC	752
Db	557	---SergLYthrThrProSerSerAlahis-----	565
OY	751	TGACCAATACAAACACCGGATCAGGTATGACAGACAGACACAACAAGATACACAGACAC	692
Db	566	----SerThrThrseryGlyArgGlyLysSerThrTrpSerArgLLeSerProGlySer	584
OY	691	AACAGATATACACTGACATGACATGACAGACGATGAGTCATGTCGTTATTACACCAACAC	632
Db	584	hrgLIuIeThr-----ThrL	589
OY	631	TCACITTTCTTCAATATTCACACCATTTCTTAGTCAGACGATACAGGTGCTATCATCCCTTC	572
Db	589	euproGlySerThrThrThrProGlyLeuSerGlyAlaSerThrThrPheTySerSerP	609
OY	571	CATAACCAACATGAAAAACACACAGTTCCAGGTGCGGAAATCCGATGCGTGTGGAGAC	512
Db	609	roArgSerProThrThrThrLeuSerProAlaSerMetThrserLeuGlyValGlyLu-	628
OY	511	TGATCATCTGATTCACACTCTGTAA GTTGAACCATTTGACGACGATTCACATTTGCAATAT	452
Db	629	-----GluSerThrThrSerArgS	635
OY	451	CGAATCCACTAGTGTGTGTGCGAATGGTGTCTCTGTGCGGCCGACATACAAGAGGCA	392
Db	635	erGlnProGlySerThr-----HisSerThv	644
OY	391	TGTATCTGTGTGCCACCTTAACTCCAGACGGTGTGTATGATGATGATGCTGTGTGCCAT	332
Db	644	alSerProAlaserThrThrThrProGly-----	653
OY	331	TCTTCAGTGTGAAGAATAAACCTGAAATATAGACCAAGATATATGACATTAAATCAT-	273
Db	654	-LeuSerGluGlnSerThrThr-----Valtyrs	663
OY	272	--GCAAAACCCACTGCATATCAAAAACACAAACGACAGCAGCAGCAAACTGTCTGTGAAA	215
Db	663	erSerSerProGlySerThrGluThr-----ThrVal-	673
OY	214	TCAATCGATTACATTAGACAAACAATGTGTTCAGTCGATCTCTTAACCCGAATCACTGGCA	155
Db	674	-----PheProArgSerThrThrThrSerValArgG	684
OY	154	AMTGAGACAAACA-----CAATGACAGCGGTGAGTACGCTGTGTGGAGACTA	104
Db	684	IyGluGlnProThrThrPheHisSerArgProAlaSerThrHisThrThrLeuPheThirg	704
OY	103	GTGAASCTGTACACATGTGTCTCTGTCTGTCTGTAAAGATGCAGTAAACGGATACAGTGC	44
Db	704	luAspSerThrThrserGlyLeuThrGluGlnuserThrAlaPheProGlySerProAlas	724
OY	43	GGGTGTGACAGAGTTACACCATCTAGTATCAACACAGTTGAC	3
Db	724	e-THrglnThrGlyLeuProAlaThrLeuThrThrAlaasp	737

RESULT 25		
Q9U3Y8		
ID	Q9U3Y8	PRELIMINARY;
AC	Q9U3Y8;	PRT; 559 AA.
DT	01-MAY-2000 (TEMBLrel. 13, Created)	
DT	01-MAY-2000 (TEMBLrel. 13, Last sequence update)	
DT	01-MAY-2000 (TEMBLrel. 13, Last annotation update)	
DE	STARP antigen.	
GN	STARP.	
OS	Plasmodium falciparum.	
OC	Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.	
OX	NCBI_TaxID=5833;	
RN	[1]	
RP	SEQUENCE FROM N.A.	
RC	STRAIN=FCCL1/HN;	
RA	Shan Z.X., Yu X.B., Li X.R., Ma C.L.;	
RT	"Cloning and characterization of Plasmodium falciparum FCCL1/HN isolate	
RT	STARP gene.";	
RL	Submitted (DEC-1999) to the EMBL/genbank/DBJ databases.	
DR	EMBL; AF209925; AAF21035.1, --	
QO	SEQUENCE 559 AA; 62167 MW; B5C1A99F0CA22C5D CXC64;	

**Alignment Scores:**

Pred. No.:	0.0015	length:	559
Score:	120.50	Matches:	84
Percent Similarity:	37.92%	Conservative:	51
Best Local Similarity:	23.60%	Mismatches:	166
Query Match:	4.52%	Indels:	115
DB:	5	Gaps:	17

US-10-020-441-1 (1-1477) x Q9U3Y8 (1-559)

QY	1384	TACACACCCAGTACTGACCTTGAAGGATGGGAGNAGAGTCCACACCTGTACCTCG	1325
DB	6	TyrIysThrIallePheThrIeuSerIleTyr-----ThrThrIeuLeuYr	21
QY	1324	AGCGGATACCTGTCGGAATTCAGACTGCTTTCACACATATAGATTGCCATTGGTCGC	1265
DB	22	Ser-----AsnYsaenLeuYsYs-----	28
QY	1264	ATGTGCGATTCCATTATYACTGCTCGACCCCTGTGTGTCAAACAAACACACACAA	1205
DB	29	-----AsnPheYrTyraenAsenAsnLeu-----SerThyTyraVal	41
QY	1204	GTAACAGTACAAAC-----	1190
DB	42	IleYsThsaaenAlyPheIeuSerGluTyryGlnIserAsnPhenleuGlyGlyTyriser	61
QY	1189	GTAACAGTACAGATAGAGAAACAAACAAACACGAGAAACAAACAAACACACAGTGAAT	1130
DB	62	AlaIalleuYsbueValaenSerYblyserGlyThraenValaenThrIysTyraen	81
QY	1129	TTCCGAAATATCATCTGGCACATATCCACCT-----TTA	1090
DB	82	SerGluasnThrAsnThrAsnAsnAsnIleProGluSerSerThrTyThrAsnThr	101
QY	1093	CACATACCACACCAATATGACACATATACATACATACATGACATAATTCAGACATCAAG	1034
DB	102	ArgIeuProIalleAsnSerThrThrSerThr--ThryValThrAsnAsnLe	121
QY	1033	TCACCTTCATCTCACACTCTCCAAACACATATCTMAACACATTTGACCCCTACTTCAT	974
DB	121	YsthrAsnIleIeuIeuThrGlyAsnAsnSerThrThrIleasn-----ThraAsn	138
QY	973	GTAACAAACACACTTCATGCTTTGGCATAGATATTTTCATGAGTCTTCTCCTCCAAAGT	914
DB	138	erThnGluasnThrSer-----	143
QY	913	GATCACAATCATCCATACATCAATTCATCCATCAATCAATCAATCAATCAATCAATCAAT	854
DB	144	-----AlaThrIyelysValThrGluasnVal-----IleThraAsn	156
QY	853	ATCAGTCACCTCACCTCACCAACACATGACCGAAGTGAATCACTCTGCTACTGAAC	794

```

Db 156 InileuThrGly-----AsnAsnAsnThrThrThrAsnThrSerThrThrGlu- 172
Qy 793 TGTCAAAAGTCACAAACACAGTGTATTATCAAGGCTGCGCATACACAAACAGG 734
Db 173 -----HisAsnAsnAsn-----IleAsnThrAsnThrAsnSerThrAspa 186
Qy 733 GATCAGTAGACAGACAGACACAAACAGATACACAGACAGAAATAGTACTGAC 674
Db 186 snSerAsnThrAsnThrAsnLeuThrAspAsnThrSerThr-ThrIleAsnLeuThrAsp 205
Qy 673 ATCTGACTGACACTGACGATGCTGTTATTACCAAACTCACTTCTTCATATTC 614
Db 206 AsnIleAsnThrThrGlnAsnLeuThrThrSerThrAsnThrThrValSer----- 223
Qy 613 CACCAATTTCTAGTACGTCAGCGATCGCTATTCAT-----CCCTTCATACCAACAA 560
Db 224 -----ThrAspAsnAsnAsnIleAsnThrIleAsnThrIleAspAsn-AsnAs 238
Qy 559 T-----GAAACACACAGTTCACAGTG 539
Db 238 nThrAspIleAsnSerThrAspAsnThrAsnThrIleAsnThrIleAsnThrAspAsn 258
Qy 538 GCGGATATCCGATGCGCTGTGGCAGACTGATCACTGATTCACCTCTGTAAGTTGACCA 479
Db 258 nThrAspIleAsn-----AlaThrAspAsnAsnAsnIleThrThrThrAspAs 275
Qy 478 TTTGAGCCGACATTCAATTGCAATATGCAATCCACTGAG 439
Db 275 nThr-----AsnThrAsnValIleSerThrAsp 284

```

## RESULT 26

Q8WMQ4 PRELIMINARY; PRT; 1349 AA.

```

ID Q8WMQ4.      PRELIMINARY;      PRT;      1349 AA.
AC Q8WMQ4.
DT 01-MAR-2002 (TREMBLrel. 20, Created)
DT 01-MAR-2002 (TREMBLrel. 20, Last sequence update)
DT 01-MAR-2002 (TREMBLrel. 20, Last annotation update)
DE Mucin 5 (Fragment).
GN MUC5AC.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21426417; PubMed=11535137;
RA Escande F., Aubert J.P., Porchet N., Buisine M.P.;
RT "Human mucin gene MUC5AC: organization of its 5'-region and central
RT repetitive region.";
RL Biochem. J. 358:763-772(2001).
DR EMBL; AJ298318; CAC83675.1; -.
FT NON_TER 1
FT NON_TER 1349
SQ SEQUENCE 1349 AA; 135600 MW; 4DC3C1544F15E8BA CRC64;

```

## Alignment Scores:

```

Pred. No.: 0.00176      Length: 1349
Score: 120.50      Matches: 77
Percent Similarity: 35.96%      Conservative: 28
Best Local Similarity: 26.37%      Mismatches: 126
Query Match: 4.52%      Indels: 61
DB: 4      Gaps: 8

```

```

US-10-020-441-1 (1-1477) x Q8WMQ4 (1-1349)
Qy 1197 ACACACAGTACGATGATGAGAAAGAAACAAACACACGGAACAAACAAACACAC 1138
Db 765 ThrSerThrIleSerAlaProThrThrSerThrThrSerThrProGlnThrSerThr--- 783
Qy 1137 AAGTAAATTCGAAATATCACTGCGACAAATATCCACCTTTACACTACCAACACACA 1078
Db 784 -----IleSerSerProThrThrSerThr-----Thr-ProThrProGln 796

```

```

Qy 1077 ATAGTCACATCAATCAATGACATGTCACATATTTCCAGCATCAAGTCACTTCATCTCAC 1018
Db 796 nThrSerThrThrSerSerProThrThrSerThrThrSerIleAlaProThrThrSerThrTh 816
Qy 1017 CACTCTCCAAACACATCACTCAAAACATTGACCCCTTACTTCATGTATACAAACACTTC 958
Db 816 rSerIleAlaProThrThrSerThrThrSerThrProGlnThrSerIleSerSer----- 833
Qy 957 CATGCTTTGCAATGATATTTATGATGCTCTCTCCCTCCAGTGCTCAATCATCCAT 898
Db 834 -----AlaProThrSerSerThrThrSerAlaPr 843
Qy 897 CACATCAATTCATCCCATCCATCCATCTACCTCACTCACTCACTCACTCACTCACTCAC 838
Db 843 oThrIleSerThrIleSerAlaProThrThrSerThrThrSerIleSerIleSerThrSerTh 863
Qy 837 TCACCAACACAAATGACCGAGATGAAATCACTCTGCTACTGAACTGTACAAAGTCATCC 778
Db 863 r-----ThrSerProProThrSerSerThrSerThrPr 875
Qy 777 AAACACAGTGTATTATCAAGGCTGTGACCAATATCAACACCGGATGAGTAGACAGAC 718
Db 875 oGlnThrSerIleThrSerIleAlaIleThrSerSerThrThrSerIleSerGlyThrThrPr 895
Qy 717 AGACACACAAACAGATACACAGACACACAGATAATGACTGACATGACATGACAGACTG 658
Db 895 oSerProValProThrThrSerThrThrIleSerValSerIleSer-----Thr 911
Qy 657 AGTCATGCTGTTATTATCAACCAACATCTCTTCAATATTCACCAATCTTACGTGA 598
Db 912 SerHisValSerValSerIleThrHisSerGlnProValThrArgAspCysHisPro 931
Qy 597 CGGATCAG-----GTCTTATCATCCCTTCC----- 571
Db 932 ArgCysThrTrpThrIleTrpIleAspValAspPheProSerProGlyProHisGlyGly 951
Qy 570 -----ATAACCAACAAATGAAACACACAGTTCACGAGGCGG 535
Db 952 AspIleGluThrIleThrAsnAsnIleIleArgSerGlyGlyIleCysArgArgProGlu 971
Qy 534 GATATCCAGATGCTGTGGCAGACTGATGATCACTGTGATTCACCTGTGTAAGTACCAATTG 475
Db 972 GlnIleThrArgGluGlnCysArgIleGluIleSerHisProGluValSerIleGlnHisIleu 991
Qy 474 AGCCGACATTCATTCGAATATGCAATTCACCTGACGTGCTGTGCGATGTTGCTCTCTG 415
Db 992 GlyIleValIleGlnCysSerArgGlu-----Glu-GlyLeuValCy 1005
Qy 414 TCGTCGCAATATACAGAGGCGATGTAT 387
Db 1005 snArgAsnGlnAspGlnGlnGlyProPhe 1014

```

## RESULT 27

Q96RZ5 PRELIMINARY; PRT; 242 AA.

```

ID Q96RZ5.      PRELIMINARY;      PRT;      242 AA.
AC Q96RZ5.
DT 01-DEC-2001 (TREMBLrel. 19, Created)
DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)
DT 01-MAR-2002 (TREMBLrel. 20, Last annotation update)
DE Putative mast cell MCP-7-like II trypase (MMP-7-like-2).
GN MCP-7L.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21096910; PubMed=1157797;
RA Daniels R.J., Peden J.F., Lloyd C., Horsley S.W., Clark K.,
RA Tufarelli C., Kearney L., Buckle V.J., Doggett N.A., Flint J.,
RA Higgs D.R.;

```

"Sequence, structure and pathology of the fully annotated terminal 2

RT Mb of the short arm of human chromosome 16.  
RL Hum. Mol. Genet. 10:339-352(2001).  
RL EMBL; AB006466; AA61272.1.  
DR InterPro: IPR001254; Ser. protease\_Try.  
DR Pfam; PF00089; trypsin\_1.  
DR PROSITE; PS50240; TRYPsin\_DOM; 1.  
DR PROSITE; PS00134; TRYPsin\_HIS; UNKNOWN\_1.  
DR PROSITE; PS00135; TRYPsin\_SER; UNKNOWN\_1.  
KW Hydroxylase; Serine protease.  
SQ SEQUENCE 242 AA; 26583 MW; 6E72DBA3EDEA2F1 CRC64;

## Alignment Scores:

Pred. No.:	0.00146	Length:	242
Score:	120.00	Matches:	37
Percent Similarity:	36.09%	Conservative:	24
Best Local Similarity:	21.89%	Mismatches:	44
Query Match:	4.47%	Indels:	64
DB:	4	Gaps:	3

US-10-020-441-1 (1-1477) x Q96R25 (1-242)

```
QY 91 TGTACAGSTTCAGTCTGCTCAACGAGACGATCTACAGCTGTCATTGTT----- 144
DB 66 CysGlyGlySerLeuIleHisProGlnTrpValLeuThrAlaAlaHisCysValGluPro 85
QY 145 --TGTCTACCATTCGACGATTCGCGGTAAAGATTCGATGAACACATTTGTTGCTCAA 201
DB 86 AspIleLysAspLeuAlaIleuLysArgValGlnLeuArgGlnHisLeuIleuYrtyrGln 105
QY 202 TGTAAATCGATTGATTTCACAGACAGTGTTCGCGTGTGCGTGTGTGTGTGTATGC 261
DB 106 AspGln----- 107
QY 262 AGTGGTTTGCATAGATTTTAATGCTATATATCTTGCTTATTTTCAGTTTCATTCTC 321
DB 108 -----LeuLeuProValSerArgIleIleValHisProGlnPheTyr 121
QY 322 ACATGAGGAATGCGGACCAACAGGATTCATCAACACCGTCTGAGTTAAGTGCGCA 381
DB 122 I1e1IeGln----- 124
QY 382 CCAGATACATGCCCTCTTGTATGTGCGACACAGAGAGACCAATCGCACAGACATC 441
DB 124 ----- 124
QY 442 AGTGATTCGATTCGATTCGATTCGATTCGATTCGATTCGATTCGATTCGATTCGATTC 501
DB 125 ThrGlyAlaAspIleAlaLeuLeuGluLeuGluProValAsnIleSerSerHisIle 144
QY 502 AGAGTGCATGCTGCGCACAGCAGATTCGCCGCGCACCTGGAACCTGGTGTTCATT 561
DB 145 HisThrValThrLeuProAlaIleSerGluThrPheProGlyMetProCysTrpVal 164
QY 562 GTTGGTTATGAAAGGATGATTAACGAC 588
DB 165 ThrGlyTyrGlyAspValAspAsn 173
RESULT 28
Q8T138 PRELIMINARY; PRT; 740 AA.
AC Q8T138;
DT 01-JUN-2002 (TREMBLrel. 21, Created)
DT 01-JUN-2002 (TREMBLrel. 21, Last sequence update)
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE Hypothetical 84.5 kDa protein.
OS Dictyostelium discoideum (Slime mold).
OC Eukaryota; Mycetozoa; Dictyostelida; Dictyostelium.
OX NCBI_TaxID=44689;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=AX4;
RA Glocker G., Eichinger L., Szafianski K., Pachebat J., Dear P.,
  Lehmann R., Baumgart C., Parra G., April J.F., Guigo R., Kumpf K.,
```

RA Tunggal B., Cox E., Quail M.A., Platzer M., Rosenthal A., Noegel A.A.;  
RT "Sequence and Analysis of Chromosome 2 of Dictyostelium."  
RL Submitted (APR-2002) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AC117075; AAM09353.1.  
KW Hypothetical protein.  
SQ SEQUENCE 740 AA; 84487 MW; 7AF5A04B780394C5 CRC64;

## Alignment Scores:

Pred. No.:	0.00178	Length:	740
Score:	120.00	Matches:	50
Percent Similarity:	38.39%	Conservative:	36
Best Local Similarity:	22.32%	Mismatches:	65
Query Match:	4.50%	Indels:	73
DB:	5	Gaps:	9

US-10-020-441-1 (1-1477) x Q8T138 (1-740)

```
QY 1220 AACAAACCAACACAGTACAGTACACAGTACAGTACAGTACAGTACAGTACAGTACAGT 1161
DB 186 AsnAsnAsnAsnAsnAsnIle-AsnAsnPheHis-----AsnAsnI 201
QY 1160 CAACGGAACAAACAAACACACAGTACAGTACAGTACAGTACAGTACAGTACAGTACAGT 1109
DB 201 eAsnAsnAsnPheAspAsnAsnSerAsnIleAsnThrIleAsnThrHisProAsnAs 221
QY 1108 -----AATATCCACCT----- 1097
DB 221 nThrAsnAsnSerSerSerAsnIleAsnGlnIleGlnSerAsnIleTyrProGlnGlnAs 241
QY 1096 -TTACACTACCA-----CACACAAATGTCACATA----- 1066
DB 241 nLeuHisAsnProPheLeuProIleHisAsnAsnAsn-AsnIleAsnAsnAsnIleA 261
QY 1065 --CAATCAATGTCGATATATTCAGCAGTACAGTACAGTACAGTACAGTACAGTACAGTAC 1008
DB 261 sNlyAsnAsnAsnAsnIlyrAsnAsnAsnAsnAsnAsnAsnIyAsnAsnAsnAsnAsnA 281
QY 1007 ACACATACATCAACACATTCAGCCCTTCTGATGATACACACACATTCATCGTTGC 948
DB 281 sAsnAsnAsnAsnAsnAsnPro----- 289
QY 947 ATGATATTTTCATGAGCTCTCTCTCTCTCAAGTGTCTCATTCATCCATCCATCCATCCAT 888
DB 290 -----GlnLeuPheSerMet-GluProProAsnGlyPheAsnHisSerThrThrAspAsn 307
QY 887 -----TTCATCCCATCAATCACT----- 867
DB 308 ThrSerSerValSerSerValProSerAsnIlySlySerSerIleThrIlyGlnIlyS 327
QY 866 -----CAACTCACTCACTCACTCACTCACTCACTCACTCACTCACTCACTCACTCACT 825
DB 328 SerIlyProLeuThrIleGlnGlnGlnGlnIlySlySerAsnIlyHisGlnGlnPro 347
QY 824 GACCGAGATGCAAAATCACTCTCTCTCTCTGAACTGTACAAAGTCAATCCAAACACATGTT 765
DB 348 AsnGlnAsnSerGlnHis-----LeuGlnSerIlySlySerSerProIle 362
QY 764 ATTATCAAGGCTCTGACCAATACCAACACCGGATCGAGTACAGACAGACACAAACA 705
DB 363 LeuIleSerSerProLeuAsnSerGlnGlnAsnSerSerProGlnProSerProThr 382
QY 704 GATACA 699
DB 383 GlnSer 384
RESULT 29
Q90WD8 PRELIMINARY; PRT; 974 AA.
AC Q90WD8;
DT 01-DEC-2001 (TREMBLrel. 19, Created)
DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE Oviductin.
```

OS Bufo japonicus (Japanese toad).  
 OC Euryptera; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Amphibia; Batrachia; Anura; Neobatrachia; Bufonidae; Bufonidae; Bufo.  
 NCBI\_Taxid=8387;  
 [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=OVIDUCTAL PARS RECTA;  
 RA Hiyoishi M., Takamune K., Mita K., Kubo H., Sugimoto Y., Katagiri C.,  
 RT "Oviductin, the oviductal protease that mediates gamete interaction by  
 RT affecting the vitelline envelope in Bufo japonicus: Its molecular  
 RT cloning and analyses of expression and post-translational  
 RT activation."  
 RL Submitted (Aug-2001) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AB070367; BAB63372.1; -  
 DR MEROPS; S01.240; -  
 DR InterPro; IPR000859; Ser\_protease\_Try.  
 DR Pfam; PF00431; CUB; 2.  
 DR Pfam; PF00089; trypsin; 2.  
 DR PROSITE; PS01180; CUB; 2.  
 DR PROSITE; PS0240; TRYPsin DOM; 2.  
 DR PROSITE; PS00134; TRYPsin\_HIS; UNKNOWN 1.  
 DR PROSITE; PS00135; TRYPsin\_SER; UNKNOWN 1.  
 KM Hydrolyse; Serine protease.  
 SQ SEQUENCE 974 AA; 107647 MW; F19705A47046553 CRC64;  
 Alignment Scores:  
 Pred. No.: 0.00268 Length: 974  
 Score: 118.50 Matches: 115  
 Percent Similarity: 33.27% Conservative: 70  
 Best Local Similarity: 20.68% Mismatches: 144  
 Query Match: 4.41% Indels: 227  
 DB: 13 Gaps: 32  
 US-10-020-441-1 (1-1477) x Q90WD8 (1-974)  
 QY 49 GAATCCCGTTCATCGCATTTCTTAAGCAGAG--AGAACATGTGTACAGTTCACATA 105  
 DB 60 GllserProtrpMetValSerleuLysArgSerGlyLysHisPheCysGlyThrIle 79  
 QY 106 GTCTCAAGAGAGCAGTACTCAGCTGCTCATTTGTGTTGCTCACCATTGCCAGTATT 165  
 DB 80 lIeserAapLysrTyValleuThrIlaIaHisCysValleuGlu----- 94  
 QY 166 CGGTAAGAGATCGACTGAACACATTGTGCTCAATGTATCATGATTTCACAGACA 225  
 DB 94 ----- 94  
 QY 226 GTGTTGGGTGCGTGGCGTTGTGTGTGTATGACAGTGGTTGCATGATTTTAATG 285  
 DB 95 -----LysAsnPhgGlu----- 98  
 QY 286 TCTATATCTGGTCTTATTTCAGTTTCATTTCACACTGAGGAATGGCACAACA 345  
 DB 99 -----PheGlnValSer-----ValSerIleGlyAspHisAsp 109  
 QY 346 GGCATTCATCAACGACGTCGAGTTAAGTGA-----CCA 384  
 DB 110 PheAlaValTyrgLysSerGlnGlnPheAlaIleLysSerValPheLysHisPro 129  
 QY 385 GGATCATGACCTCTTGTATGTGCGCAGACAGAGAGACCAATGCGACAGACACTCAGT 444  
 DB 130 AsnPheLysProSer-----ArgProPhe----- 137  
 QY 445 GGATTGATTTGCAATGTATGCTGGCTCAATGTGTCACTTACAGAGTGAATCA 504  
 DB 138 AsnTyraPheLeuAlaIleLeuGlnLeuValGlnSerIleThrPheAspLysAspIleGln 157  
 QY 505 GTGATCACTGCGCAGACATCGATATCCCGCAGACCTGACCTGAGTGTTCATTGT 564  
 DB 158 ProAlaCysLeuProSerProAspAspValPheProthnGlyThrLeuCysMetAlaLeu 177  
 QY 565 GGTATGGAAG--GATGATTAACGACGTCGTCGTCACGTAAAGTGTGAATATTG 621

DB 178 GlyTrpGlyArgLeuGlnGlnAsnGlyArgLeuProSerSer-----Leu 192  
 QY 622 AAGAAAGTGAAGTGTGTGGAATAAAGACATGACTGCTGTCAGTCAAGATGTCATCA 681  
 DB 193 GlnLysVal----- 195  
 QY 682 GTTATCTGTGTGTCGTGTGTATCTGTTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 741  
 DB 196 ValLeuProLeuIleGlyTyraArgArgCysLeuSerIle-----MetGln 210  
 QY 742 GTATTGTCAAGCCTGTATATAACAACCTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 801  
 DB 211 ThrValAspArgArgLeuAlaPheGlnThrVal-----Val 222  
 QY 802 GCAGAGTATTTCCA-----TTCGGTCATTGGTGTGG-----TGAGTGAAGTGAACG 849  
 DB 223 CysAlaGlyPheProGlnGlyGlyLysAspAlaCysGlnGlyAspSerGlyLysProPhe 242  
 QY 850 TGATGTGAGTGAAGTGTGAGTGTGATGATGATGATGATGATGATGATGATGATGATGAT 892  
 DB 243 LeuCysGlnArgSerGlnGlyArg-TyrValLeuValGlyValThrSerTrpGlyLeuG 262  
 QY 893 -----ATGTGATGGATGAT-----TGACACCACTTGACAGAGACA----- 928  
 DB 262 CysAlaArgLysTrpAlaAspAsnIleLeuAspProValGlnSerLysGlySerProG 282  
 QY 929 -----AAGCTCATGAATATTTATGCAACAGATGGAAGTGTGTGT 969  
 DB 282 yValPheThrAspIleGlnAlaGlyLeuAsnTrpLeuSerGlu----- 296  
 QY 970 GTACATGAAGTAGGCGGTCAATGTGTGTGAGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1029  
 DB 297 -----AsnLeuAsnGlnAspLysProAspPheProThrTy----- 308  
 QY 1030 GTGACTGATCGTCGATATATATGTCACATGTGAT----- 1063  
 DB 309 -----GlnValGlnCysSerThrAsnAspGlyIleGlyLysGlyThrThrG 324  
 QY 1064 ----- 1071  
 DB 324 yGlnIleLeuLeuProThrGlyTyLysTyLysTySerAsnAsnGlnLysCysIleTr 344  
 QY 1072 GACTATGTTGTGTGCGGTAGTGAAGGTGATATGTCACAGTGTATTTTGAAT 1131  
 DB 344 PThrIle-----IleValProArgGly-----Lys 352  
 QY 1132 TCAGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1188  
 DB 352 shIleLeuLeuThrPheLysSerPheAsnValGlnCysAspTyLysSerCysAspLeuAs 372  
 QY 1189 CGTTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1248  
 DB 372 pTyLeuValIleTySerAlaLeuGlyArgLeu-IleGlyLys----- 386  
 QY 1249 AATGGAATCGGACATGAGGACATGCAATGCAATTCATATGTGTGAAGAGGTCAGAATTT 1308  
 DB 387 --PheCysGlyAspValSerProArgProLeuLeu-----IleA 399  
 QY 1309 CGGACAGTTAACCGCTCCAGGTGACAGTGTGAGCTCTCTCC--CATCCCTTCAAGG 1365  
 DB 399 lAspAla-----SerIleThrLeuLysPheIleSerAspPheHisGlyTyLysTr 416  
 QY 1366 TCAGTACTGCGGTGCGTATCATATGTTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1422  
 DB 416 hrgLysPheSerLeuPheTyrgLysAlaValGluPro--AspThrTyProAspSerAspC 435  
 QY 1423 TGTGAGTATGCAAGTGTGCTAGAA----- 1448  
 DB 435 yGlySerValAlaValIlePheGlnGlnGlyGlyIleGlnThrMetAsnHisProHisL 455  
 QY 1449 -----TGT--TGATTTGTACGCTCA 1469

Db 455 eutYrSerSerHisAlaAsnCysGlnTrpValValHisSerPro 469

RESULT 30

0960L8 PRELIMINARY; PRT; 270 AA.

AC 0960L8

DT 01-DEC-2001 (TREMBLrel. 19, Created)

DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)

DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)

DE Similar to elastase 3, pancreatic (protease E).

OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.

OX NCBI\_TaxID=9606;

RN [1]

RP SEQUENCE FROM N.A.

RC TISSUE=PANCREAS;

RA Strausberg R.;

RL Submitted (MAY-2001) to the EMBL/GenBank/DBJ databases.

DR EMBL; BC008383; AA08383.1; -.

DR MEROPS; S01.154; -.

DR InterPro; IPR001254; Ser\_protease\_Try.

DR Pfam; PF00089; trypsin\_1.

DR PROSITE; PS0240; TRYPsin\_DOM; 1.

DR PROSITE; PS00134; TRYPsin\_HIS; UNKNOWN\_1.

DR PROSITE; PS00135; TRYPsin\_SER; UNKNOWN\_1.

KW Hydrolase; Protease; Serine protease.

SQ SEQUENCE 270 AA; 29446 MW; 946DDDA64A102E CRC64;

Alignment Scores:

Pred. No.:	Score:	Length:	Matches:	Conservative:	Mismatches:	Indels:	Gaps:
US-10-020-441-1 (1-1477) x 0960L8 (1-270)	0.00241	270	118.00	43	26	60	4
Percent Similarity:	34.33%						
Best Local Similarity:	21.39%						
Query Match:	4.39%						

DB:

US-10-020-441-1 (1-1477) x 0960L8 (1-270)

QY 25 GGTGAACCTGTCACACCGCACT----- 48

DB 19 GlyProProSerSerHisSerSerSerArgValValHisGlyAlaValProTyr 38

QY 49 GAATCCCGCTGATCGCATTTTAAACGACGAGAACCA-----ATGCT 93

DB 39 SerTrpProTyrGlnValSerLeuGlnTyrGlnValSerGlySerPheTyrHisThrCys 58

QY 94 ACAAGTCTCTAGTCTCAACGACGAGACTACTCAGCTGGTGCATTTGGTCTCACA 153

DB 59 GlyGlySerLeuIleAlaProAspTrpValValThrAlaGlyHisCysIleSerArgAsp 78

QY 154 TTGGCAGTATTCGGGTAAAGATGACCTGAAACACATTTGGTCAATGATATCATTTG 213

DB 79 Leu----- 79

QY 214 ATTTCACAGACAGTGTTCGTGCGTGCCTGTTGTGTGTGTATGACAGTGGTTTGA 273

DB 79 ----- 79

QY 274 TAGATTTTATGCTATTAATCTGCTTATTTTCAAGTTTCAATTTTCTACACTGAGAAAT 333

DB 80 -----ThryrGlnValValLeuGlyGlyTyrAsnLeuAlaValIysGln 94

QY 334 GGGACCAACAAAGGATCATCCACACGCTCGAGTTAAGGTAGCGACAGATACATG 393

DB 95 GlyProGlnValIleProIleAsnSerGlnLeuPheValHisProLeuTyrPaa 114

QY 394 CCTCTTGTATGTGGCAGCAGACAGAGACCAATGACACAGACTCACTGAGTTGAT 453

DB 115 ArgSerCysValAlaCys-----GlyAsnAsp 123

QY 454 ATTGCATTTGATGCTGGCTCAATGTGTCATTAACAGATGGAAATCAAGATGATCAAT 513

Db 124 IleAlaLeuIleLysLeuSerArgSerAlaGlnLeuGlyAspAlaValGlnLeuAlaSer 143

QY 514 CTGCCACAGCCATCGATATCCCGCACTGGAACCTGGTGTTCATTTGTTATGGA 573

DB 144 LeuProProAlaGlyAspIleLeuProAsnLysThrProCysTyrIleThrGlyTyrGly 163

QY 574 AGG 576

DB 164 Arg 164

RESULT 31

090U18 PRELIMINARY; PRT; 166 AA.

ID 090U18

AC 090U18

DT 01-MAY-2000 (TREMBLrel. 13, Created)

DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)

DT 01-MAR-2002 (TREMBLrel. 20, Last annotation update)

DE Mast cell mGCP-7-like I protein (Fragment).

OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.

OX NCBI\_TaxID=9606;

RN [1]

RP SEQUENCE FROM N.A.

RE MEDLINE=99121069; PubMed=9920877;

RA Pallao M., Fejzo M.S., Shayesteh L., Blount J.L., Caughey G.H.,

RT "Characterization of genes encoding known and novel human mast cell

RT tryptases on chromosome 16p13.3."

RL J. Biol. Chem. 274.3355-3362(1999).

CC - SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE

CC TRYPSIN FAMILY.

DR EMBL; AF099147; AADI7861.1; -.

DR HSSP; P20231; IAAO.

DR MEROPS; S01.054; -.

DR InterPro; IPR001314; Chymotrypsin.

DR InterPro; IPR001254; Ser\_protease\_Try.

DR Pfam; PF00089; trypsin\_1.

DR PRINTS; PR00722; CHYMOTRYPSIN.

DR SMART; SM00020; Tryp\_SPC; 1.

DR PROSITE; PS0240; TRYPsin\_DOM; 1.

DR PROSITE; PS00134; TRYPsin\_HIS; UNKNOWN\_1.

KW Hydrolase; Serine protease.

FT NON\_TER 166

SQ SEQUENCE 166 AA; 18433 MW; A7AF897A6423E7D5 CRC64;

Alignment Scores:

Pred. No.:	Score:	Length:	Matches:	Conservative:	Mismatches:	Indels:	Gaps:
US-10-020-441-1 (1-1477) x 090U18 (1-166)	0.00282	166	117.00	36	25	44	3
Percent Similarity:	36.09%						
Best Local Similarity:	21.30%						
Query Match:	4.36%						

DB:

US-10-020-441-1 (1-1477) x 090U18 (1-166)

QY 91 TGTACAGTTCACATGCTCAACGACGAGACTACTCAGCTGATCTGTT----- 144

DB 59 CysGlyGlySerLeuIleHisProGlnTrpValLeuThrIleAlaHisCysMetGluPro 78

QY 145 ---TGTCAACATTCGCGAGTATTTGGGTAAAGATTCAGTGAACACATTTGCTCA 201

DB 79 AspIleLysAspLeuAlaLeuArgValGlnLeuArgGlnHisLeuTyrTyrGln 98

QY 202 TGTATTCATTTATTTACAGACAGATGTTGCGTGGCTGCTTGTGTGTGTATGTC 261

DB 99 AspGln----- 100

QY 262 AGTGGTTTGCATAGATTTTATGCTATATCTGCTTATTTTCAAGTTTCAATTTCTC 321

DB 101 -----LeuLeuProValSerArgIleIleValHisProGlnPheTyr 114

QY 322 ACATGAGGATTCGGACCAACAGAGCATTCATCCACCAACGCTTGAGTTAAGTGGCA 381



```

Db 115 IleIleGln----- 117
Qy 382 CCAGATACATGCGCTTGTATGTGCGACAGACAGACCAATCGCACAGACATC 441
Db 117 ----- 117
Qy 442 AGTGATTCGATATTGCAATTGTAATGCTGCTCAATGTCATTAACAGATGGAATC 501
Db 118 ThrGlyAlaSerPheLeuLeuGluLeuGluProValAsnIleSerHisIle 137
Qy 502 AGAGGATCAGCTGCGACAGCCATCGATATCCGCCCTCGGAACAGCTGTTTCAT 561
Db 138 HisThrValThrLeuProProAlaSerGluThrPheProGlyMetProCysTrpVal 157
Qy 562 GTTGCTTATGGAAGGATGATTAACGAC 588
Db 158 ThrGlyTrpGlyAlaSerValAspAsnAsn 166

RESULT 32
Q9H8D9 PRELIMINARY; PRT; 858 AA.
ID Q9H8D9 AC Q9H8D9:
DT 01-MAR-2001 (TREMBLrel. 16, Created)
DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)
DT 01-JUN-2002 (TREMBLrel. 17, Last annotation update)
DE CDNA FLJ13725 fis, clone PLACE300009, weakly similar to DNA-directed
DE RNA polymerase II largest subunit (EC 2.7.7.6).
OS Homo sapiens (Human)
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=PLACENTA;
RA Isegai T., Oca T., Hayashi K., Sugiyama T., Otsuki T., Suzuki Y.,
RA Nishikawa T., Negai K., Sugano S., Takahashi-Fujii A., Hara H.,
RA Tanase T., Nomura Y., Togiya S., Komai F., Hara R., Takeuchi K.,
RA Arita M., Nabekura T., Ishii S., Kawai Y., Saito K., Yamamoto J.,
RA Wakamatsu A., Nakamura Y., Nagahari K., Masuo Y., Oshima A.,
RT "NEDO human cDNA sequencing project.";
RL Submitted (AUG-2000) to the EMBL/Genbank/DBJ databases.
DR EMBL; AK023787; BAB14678.1;
DR InterPro; IPR002965; P rich extensn.
DR PRINTS; PRO1217; PRICHEXTENSN.
SQ SEQUENCE 858 AA; 91130 MW; 46762a89AB847415 CRC64;

Alignment Scores:
Pred. No.: 0.00425 Length: 858
Score: 116.50 Matches: 96
Percent Similarity: 30.20% Conservative: 26
Best Local Similarity: 23.76% Mismatches: 118
Query Match: 4.37% Indels: 164
DB: Gaps: 22

US-10-020-441-1 (1-1477) x Q9H8D9 (1-858)
Qy 1413 GGAAGGTTGGAGAGTGACACCATGTGATCGACACGAGTACGACCTTGAAGGAT 1354
Db 149 GlyThrThrGlySerVal---ProHiserThrAspProAlaProSerAlaHisLeuAsp 167
Qy 1353 GGG-----AGGAGAGTCCACCATGTCATGTCAGCTGAGCGGGTAACTGTCG 1309
Db 168 SerValHisLysSerThrAspSerGlyProSerGluLeuProGly----- 182
Qy 1308 AAATTTCAGCTGCTTTACACATATAGATGTCATTGGTCGATGCGCATTCATT 1249
Db 183 -----ProThrHisThrThrThrGly----- 189
Qy 1248 ATAGTCGCTCGACCTGTGGTGAACAAACAAACAAACAAACAGTACAGT-----ACA 1195
Db 190 -----SerThrTyrSerAlaIleThr 196
Qy 1194 ACAACGTACAGTACGATGAGAAAGAACAAACAAACGGAACAAACAAACAAACAG 1135

```

```

Db 197 ThrThrHisSerAlaProSerProLeuThrHisThr-----ThrGlySerThrHisLys 215
Qy 1134 TGAATTGGAATTAATTCAGTGGACATATCCACCCCTTACAGCCACACAAACATA 1075
Db 216 ProIle-----IleSerThrLeuThrThrThrGlyProThrLeu 228
Qy 1074 GTCCACATACATCATGTCGACATATGATTCGAGCATCAATGAC-----TCTCCA 1024
Db 229 ---AenIleGlyProValGlnThrThrHisSerProThrHisThrMetProSerPro 247
Qy 1023 TCTCAC-----CACTTCCAAACATATCTCA-----ACACATTAACCCCTTAC 979
Db 248 ThrHisThrThrAlaSerPro-ThrHisThrSerThrSerProThrHis---ThrProth 266
Qy 978 TTCAATGACACACACATTCGATTCGATTCGATTCGATTCGATTCGATTCGATTCG 921
Db 266 IserProThrHisLysThrSerMet-----SerProPr 277
Qy 920 -----CCAAGTGTCTCAT----- 906
Db 277 OThrThrHiserProThrProSerGlyMetSerLeuValGlnThrAlaThrSerProth 297
Qy 905 -CATCCATCAATCATTCATTCATTCATTCATTCATTCATTCATTCATTCATTCATTC 867
Db 297 HiserProThrHiserProThrHiserProThrHiserProThrHiserProThrHis 317
Qy 866 -----CACTGACCTGACATGACGTCGATGACCTGACCTGACCTGACCTGACCTG 822
Db 317 OserThrSerLeuGluLeuAlaThrLeuSerSerProSerHisSerAspProThrLe 337
Qy 821 -----CGAGTGAATATCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 769
Db 337 uProGlyThrAspSerLeuProCysSerProProValSerAsnSerTyr----- 353
Qy 768 TGTATTATTCAGAGCTGTGACCAATACAAACCGGATCAGTATGACAGACACACACA 709
Db 354 -----ThrG1 355
Qy 708 AACAGATACAGACACACAGATATGACTGACATCTGACCTGACCTGACCTGACCTG 649
Db 355 nAlaAspProMetAlaProAlaGlyThrPro----- 364
Qy 648 GTTATTACCAACAACATCACTTCTTCATATTCACCATTTCTTAGT-----GA 598
Db 365 -----HisProSerProAlaHisSerSerThrGlySerProLeuThrHiserProAlaProAs 382
Qy 597 CGATCAGCGTGTATTCATCCCTTCCATTAACCAACATGAAACACACAGTTCGAGGTG 538
Db 382 pProSerGluSerThrValGlnSerLeuSerProThrProSerProProThrProAlaPr 402
Qy 537 CGGATATCCAGTGGCTGTGCGACAGCTGATCATCTGATTCATCTGTGAATTGACCAT 478
Db 402 oGlnHisSerAspLeuCys----- 408
Qy 477 TTGAGCCAGATTAATTCATATGATATGATGATGATGATGATGATGATGATGATGAT 423
Db 409 -----LeuAlaMetAlaValGlnThrProValProThrAlaAlaGlyLysSerG1 425
Qy 422 -----CTCTGTGTGTCGCGACATACATCAAGAGG 394
Db 425 yAspAspSerLeuGluGluAlaLeuGlyAlaLeuMetAlaAlaLeuMetAspTyrArgG1 445
Qy 393 CATGTATCTT 384
Db 445 yGlnPhePro 448

RESULT 33
Q8SSU1 PRELIMINARY; PRT; 1492 AA.
ID Q8SSU1 AC Q8SSU1:
DT 01-JUN-2002 (TREMBLrel. 21, Created)
DT 01-JUN-2002 (TREMBLrel. 21, Last sequence update)

```

01-JUN-2002 (TREMBlrel. 21, last annotation update)  
DE Putative guanine-nucleotide releasing factor.  
OS Dictyostelium discoideum (Slime mold).  
OC Eukaryota; Mycetozoa; Dictyostelida; Dictyostelium.  
OX NCBI\_TaxID=44689;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=AX4;  
RA Gloeckner G., Eichinger L., Szafranski K., Pachebat J., Dear P.,  
Lehmann R., Baumgart C., Parra G., April J.F., Guiso R., Kumpf K.,  
Tunggal B., Cox B., Quail M.A., Placer M., Rosenthal A., Noegel A.A.,  
RT "Sequence and Analysis of the EMBL/Genbank/DBJ databases."  
RL Submitted (MAR-2002) to the EMBL/Genbank/DBJ databases.  
DR EMBL; AC115679; AAL92624.1; -  
SQ SEQUENCE 1492 AA; 164266 MW; 5E3BD343D4A8605 CRC64;

Alignment Scores:  
Pred. No.: 0.00468 Length: 1492  
Score: 116.50 Matches: 59  
Percent Similarity: 38.33% Conservative: 33  
Best Local Similarity: 24.58% Mismatches: 74  
Query Match: 4.37% Indels: 74  
Gaps: 9

US-10-020-441-1 (1-1477) x Q8SSU1 (1-1492)

QY 1257 CATTCATTATATGTCGTCGACCCCTGTGGTGCAACAAACACACACAGTACAGT 1198  
DB 814 AenahenValillierSeramhenthrglyAenahen-AenahenAenahenSerSerly 833  
QY 1197 ACAACACGTCATGTCAGATGAGAAAGAAACAAACACGGAACAAACAAACACAC 1138  
DB 833 sasnahenanthrVal-----LysglyAenahenLysglynginAenahenLysSerSerse 851  
QY 1137 AAGTGA----- 1131  
DB 851 rserginAenahenSerProProSerAerPheserLeuSerValSerProSerProCysse 871  
QY 1130 -----TT 1129  
DB 871 rserleuthrProserProserProserProserProserProleuilegluSerSerMetph 891  
QY 1128 TCGAAATATATCACTGGACAAATATCCACCTTTACA-----CTACCCACACACA 1078  
DB 891 ethrlyrs--AspleuSerProProProProlaalaalalaalaleuilegluInlndr 910  
QY 1077 ATATGTCATATCATCATGTCACCTATATTCGACGATCAAGTCACTCTC----- 1026  
DB 910 ovala1athrthrAenahenLysSerThrAspLysSerArgLysSerleuThrPserse 930  
QY 1025 -----CATTCACGACCTCTCCAAACACATAC 1000  
DB 930 rthetlySerSerlyrProProLleserThrserginProserThrilleghinSphThrI 950  
QY 999 TCGAAACACATGACCCCTACTTCATGATACACAAACACTTCCTCATGTTGCATAGATAT 940  
DB 950 eserThrser---ProProSerThrserThrserProthrSer----- 963  
QY 939 TTCATGAGCTTCTCTCTCCATGAGTGTCTCAATCATCCCATACATCATTCATCC 880  
DB 964 -----AsnSerProProThrSerilleSerProProProProSerilleThrI 980  
QY 879 ATCCCATCTACCTCACTCACTCACTCACTCACTCACTCACTCACTCACTCACTCACT 820  
DB 980 ySerSerProProSerThrAlaely-----ThrserProProAsnAspAsnlyAs 997  
QY 819 AGATGAAATCACTCTGTCTACTGATGATGTCACAAAGTCACTCAACAGTTGTTATAT 760  
DB 997 nAenahenAenahenSerAenahenLysgluLeuThrilleSerilleProilleThrAenahenVal 1017  
QY 759 C-----MAGGCTGACCAATACAAACACCGGATGAGTACAGACACACACA 711  
DB 1017 lHnsluSerAenahenValValalaserThrProArgSerGlyArgAspLeuAlathr 1036

RESULT 34  
ID 026021 PRELIMINARY; PRT; 604 AA.  
AC 026021;  
DT 01-NOV-1996 (TREMBlrel. 01, Created)  
DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)  
DT 01-NOV-1998 (TREMBlrel. 08, Last annotation update)  
DE STAP antigen.  
OS Plasmodium falciparum.  
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.  
OX NCBI\_TaxID=5833;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=T9/96;  
RX MEDLINE=95021499; Pubmed=7935600;  
RA Fidock D.A., Bortius E., Brahimi K., Moelans I.I.M.D., Aikawa M.,  
Königs R.N.H., Cetta U., Olafsson P., Kaldoh T., Asavanich A.,  
RA "Cloning and characterization of a novel Plasmodium falciparum  
sporozoite surface antigen, STAP."  
RT Mol. Biochem. Parasitol. 64:219-232(1994).  
RL EMBL; Z26314; CAA81224.1; -  
SQ SEQUENCE 604 AA; 67051 MW; DE23D9442A509667 CRC64;

Alignment Scores:  
Pred. No.: 0.00573 Length: 604  
Score: 115.00 Matches: 83  
Percent Similarity: 37.31% Conservative: 61  
Best Local Similarity: 21.50% Mismatches: 112  
Query Match: 4.31% Indels: 130  
Gaps: 17

US-10-020-441-1 (1-1477) x Q26021 (1-604)

QY 1384 TACGACACCGAGTACGACCTTGAAAGGATGGAGAGAGTCCACCATGTCACCTG 1325  
DB 6 TyrlystrAlaallePherthrleuSerilletrp-----ThrThrleuLeuYr 21  
QY 1324 AGCGGGTAACGTGTCGGAATTCGACCTGCTTCACACATATAGGATTCGATTCGTCG 1265  
DB 22 Ser-----AsnlyAsnleuLeuYrCys----- 28  
QY 1264 ATGTGCGATTCATATATAGTCGTCGACCTGTGTGCAACAAACACACACACACA 1205  
DB 29 -----AsnPhetYrTyAsnAsnAsnleu-----SerThrYrVal 41  
QY 1204 GTACGTCACAC-----AAC 1190  
DB 42 lleyshAsnArgPheleuSergluYrInserAenPheleuGlyGlyYrYrSer 61  
QY 1189 GTACAGTACAGTACGAAAGAAACAAACACGGAACAAACAAACACACACATGAT 1130  
DB 62 AlaAlaleuLysleuValAsnSerlySylserglYThrAenValAsnThrlyeYrAsn 81  
QY 1129 TCGAAATATCACTGGACAAATATCCACCTTTACACTACCCACACACAAATAGTCCA 1070  
DB 82 SergluAenThrAenThrAenAsn-----AsnilleProgluSerSer--SerT 97  
QY 1069 CATTCATCATCATGTGACCTATATTCGACGATCAAGTCACTCTCCATCCATCCATCC 1010  
DB 97 hrTYrThrAenThrArgleuAlaAenAenSerThrThrThrSerThrThrlyrValT 117  
QY 1009 AAACATATCTCAACACATTCGACCC-----CCTACTT 977  
DB 117 hrAspAsnAsnlystrAsnilleLysleuThrGlyAsnAsnSerThrThrilleAenThrA 137  
QY 976 CATGTAACACACACTTCATGCTTGTGATGATGATATTCATGAGTCTCTCTCTCA 917  
DB 137 snSerThrgluAsnThrser-----Alathrlysl 147  
QY 916 GTGTCTCATATCCATCCATCATCATATTCATCCATCCATCCATCCATCCATCCATCC 857

```

Db 147 yvalThrluAnValIleThraEngInIleLeuThrglyAsnAsnThrThr 167
QY 856 CACATCAGCTCAGCTCAGCTCAGCAACATGACGGA----- 819
Db 167 sn-----ThserThrThrGluHisAsnAsnIleAsnThrAsnSer 184
QY 818 -----G 818
Db 184 hrGluAsnThrSerAlaThrLysLysValThrGluAsnValIleThrAsnInIleLeu 204
QY 817 ATGAAATCAGCTCTGCTACTGAGTGTCAAAAGT-----CATCAAAACAGATTGTA 764
Db 204 hrGlyAsnAsnThrThrThrAsnThrSerThrThrGluHisAsnAsn----- 221
QY 763 TTATGAAGCTCTGACCAATACACACCGGATCGGTAGACAGACAGACAAACAG 704
Db 222 --IleAsnThrAsnThrAsnSerThrAspAsnSerAsnThrAsnThrAsnLeuThrAspA 241
QY 703 ATACACAGACACAGAAATMACTGACATGCTGACGACGAGTCAATGCTGTTA 644
Db 241 AsnThrSerThr-ThrLysLysLeuThrAspAsnIleAsnThrThrGluAsnLeuThrThr 260
QY 643 TTCACCAACCAACTCACTTTCTTCATATTCCACCATTTTACGTGACGATCAGGTCGT 584
Db 261 SerThrAsnThrThrThrValSer-----ThrAspAsnAsn 273
QY 583 TATCAT-----CCCTTCATACCAACAAAT----- 559
Db 274 IleAsnThrLysProIleAsn--AsnAsnThrAspIleLysSerThrAspAsnThrAs 293
QY 558 -----GAAACACCAAGTTCCAGGTGCGGATATCCGATGCTGTGACAGATGA 509
Db 293 nThrglyThrLysGluThrAspAsnLysAsnThrAspIleLys-----AlaThrAs 310
QY 508 TCACCTGATTCCTCTGTAAGTTGACCATTTGAGCCAGCATTCACATTCGAATATGCA 449
Db 310 pAsnAsnAsnIleThrThrThrThrAspAsnThr-----AsnThrAsnValI 326
QY 448 ATCCAGCTGAG 439
Db 326 eserThrAsp 329

RESULT 35
Q8ST04 PRELIMINARY; PRT; 1461 AA.
ID Q8ST04;
AC Q8ST04;
DT 01-JUN-2002 (TRENBLrel. 21, Created)
DT 01-JUN-2002 (TRENBLrel. 21, Last sequence update)
DT 01-JUN-2002 (TRENBLrel. 21, Last annotation update)
DE Hypothetical protein.
OS Dictyostelium discoideum (Slime mold).
OC Eukaryota; Mycetozoa; Dictyostelida; Dictyostelium.
OX NCBI_TaxID=44689;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=AX4;
RA Gloeckner G., Eichinger L., Szafranski K., Pachebat J., Dear P.,
  Lehmann R., Baunsgart C., Paria G., April J.F., Guigo R., Kumpf K.,
  Ralunggal B., Cox B., Quail M.A., Platzer M., Rosenthal A., Noegel A.A.,
  "Sequence and Analysis of Chromosome 2 of Dictyostelium,"
  Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AC115681; AAL92653.1; -
KW Hypothetical protein.
SQ SEQUENCE 1461 AA; 169095 MW; A867DA194858EAS5C CRC64;

```

## Alignment Scores:

```

Pred. No.: 0.00669 Length: 1461
Score: 115.00 Matches: 59
Percent Similarity: 38.49% Conservative: 38
Best Local Similarity: 23.41% Mismatches: 93
Query Match: 4.31% Indels: 62
Gaps: 11

```

```

US-10-020-441-1 (1-1477) x Q8ST04 (1-1461)
QY 1241 CTGCAGCTGTGTGTCGAACAACCAACCAACAGATGATCAACCAAGTACAGTA 1182
Db 779 LeuAsnProLysIle-----SerAsnSerLeuAsnLysThrIleIleAsnAsnIle 796
QY 1181 CAGATGAGAAAGAACMAAACACACACGAAACAAACMAAACACACAGATTTGAA 1122
Db 797 AsnIleAsnAsnAsnAsnAsnAsnAsnAsnAsnAsnIleAsnAsnSer-AsnAsnAsn 816
QY 1121 TATC-----AACTGGCAATATATCCACCTTTACACTACCAACCAACATATGTCACA 1068
Db 816 nIleHisIleAsnHisIleHisIleHisIleHisIleHisIleHisIleHisIleHisIle 836
QY 1067 TACATTCACATGTGACATATATTCAGATGAGTCAAGTCACTTCATCTGACCACTCCAA 1008
Db 836 n-AsnAsnAsn-----AsnAsnAsnAsnAsnAsnAsnAsnAsnAsnAsnAsn 852
QY 1007 ACACATCTCAAAACACATTCAGCCCTTACTTCATGACAAACACATTCATCGTTGC 948
Db 852 snAsnAsnAsnAsnAsnAsnAsnThrProGln-----HisGlyLeuA 865
QY 947 ATAGATATTC-----ATGAGCTTTCTCTCTCCAGATGCTCAAT---CATCC 900
Db 865 snAsnIleLeuThrAlaSerLeuAsnLeuSerSer-AsnValLeuAsnSerThr 884
QY 899 ATCAATGACATTCATCCATCCCA-----TCCAAATCTCACTCACTGACCTCA 855
Db 885 AlaThrAlaAlaGAsnIleProLeuSerThrThrThrAsnIleProPro---ThrAsp 903
QY 854 CATCAGCTCAGCTCAGCTCAGCAACACACATGACCGAGATGGAATCACTGTGTA 795
Db 904 HisProIleGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 923
QY 794 CTGTCAAAAGTATCCAAACACAGTTGTATTA-----TCAAG 756
Db 924 GlnGlnGlnProGlnProGlnGlnGlnGlnLeuProGlnHisGlnHisGlnGlnGln 943
QY 755 GCTGTGACCAATATCAACACACGAGTCAAGTACAGACACACAAACAGATACAG 696
Db 943 nGlnGlnProHisGlnGlnProHisGlnHisGlnGlnGlnGlnGlnGlnGlnProHisG 963
QY 695 ACACACAGAAATATGACTGACATCTGACGTGACGACTGATGATCGTTATTCACCA 636
Db 963 nGlnPro----- 965
QY 635 CAATCACTTTCTTCATATTCACCATTTTACGTGACGATCAGGTCGTTATCATCC 576
Db 966 -----HisHisGlnGln---GlnHisHisGlnAsnAsnHisG 977
QY 575 CTTCATTAACCAACAAATGAAACACACAG 548
Db 977 nGlnHisGlnGlnHisGlnGlnHisGln 986

RESULT 36
Q9V918 PRELIMINARY; PRT; 1976 AA.
ID Q9V918;
AC Q9V918;
DT 01-MAY-2000 (TRENBLrel. 13, Created)
DT 01-MAR-2001 (TRENBLrel. 16, Last sequence update)
DT 01-JUN-2002 (TRENBLrel. 21, Last annotation update)
DE CG14470 protein.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
  Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
  OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BERKELEY;
RX MEDLINE=20196006; Pubmed=10731132;
  Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,

```

RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,  
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,  
 RA Sutton G.G., Wortman J.R., Vandeil M.D., Zhang Q., Chen L.X.,  
 RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe C.M., Pfeiffer B.D.,  
 RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,  
 RA Abell J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,  
 RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,  
 RA Beeson K.Y., Benos P.V., Bernan B.P., Bhandari D., Bolshakov S.,  
 RA Botkin C.D., Bousman M.R., Bouck J., Brockstein P., Brodtier P.,  
 RA Butke K.A., Busan D.A., Butler H., Cadieu E., Center A., Chandra I.,  
 RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,  
 RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,  
 RA Dodson K.J., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,  
 RA Fowler C.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,  
 RA Foster C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,  
 RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,  
 RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,  
 RA Hosteln D., Houston K.A., Howland T.J., Wei M.-H., Ibeagwu C.,  
 RA Jajalji M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,  
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai X.,  
 RA Liang X., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,  
 RA Liu X., Maltsev B., McIntosh T.C., McLeod M.P., McPherson D.,  
 RA Mekulov G., Mishina N.V., Mobarry C., Morris J., Moshrefi A.,  
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,  
 RA Nelson D.R., Nelson K.A., Nixon K., Nuskern D.R., Paclab J.M.,  
 RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,  
 RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,  
 RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,  
 RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,  
 RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,  
 RA Wang Z.-Y., Wasserman D.A., Weinstein G.M., Weissbach J.,  
 RA Williams S.M., Woodage T., Wortley K.C., Wu D., Yang S., Yao Q.A.,  
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang X., Zhao Q., Zheng L.,  
 RA Zhang X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,  
 RA Gibbs R.A., Myers E.W., Rubin G.W., Venter J.C.,  
 RA "The genome sequence of Drosophila melanogaster."  
 RT Science 287:2185-2195(2000).  
 CC -1 SIMILARITY: BELONGS TO THE SERPIN FAMILY.  
 DR EMBL: AE003785; AAF57301.2; -.  
 DR HSSP: P05121; 1ATC.  
 DR FLYBase: FBgn0033046; CG14470.  
 DR InterPro: IPR002965; P\_rich\_extensn.  
 DR InterPro: IPR00215; Serpin.  
 DR Pfam: PF00079; serpin; 1.  
 DR PRINTS: PR01217; PRICHEXTENS.  
 DR SMART: SM00093; SERPIN; 1.  
 DR PROSITE: PS00284; SERPIN; UNKNOWN\_1.  
 KW Serpin.  
 SQ SEQUENCE 1976 AA; 216926 MW; 61461696532918D9 CRC64;

Alignment Scores:  
 Pred. No.: 0.00706 Length: 1976  
 Score: 115.00 Matches: 70  
 Percent Similarity: 33.14% Conservative: 42  
 Best Local Similarity: 20.71% Mismatches: 91  
 Query Match: 4.31% Indels: 135  
 DB: S Gaps: 15

US-10-020-441-1 (1-1477) x Q9V918 (1-1976)

QY 1392 CCATGTGATGACACACCGATGACCTTGAAGGATGGAGAGAGTCCACCATG 1333  
 DB 319 ProLaseThrPro-----LysLysLysLysLysLysProThrArg 333

QY 1332 TCACCTGGAGCGGTACTGT-----CCGAATTCACCT 1297  
 DB 334 ThrProGlyThrGlySerSerGlySerSerLysProSerLysProLysAlaGlyPro 353

QY 1296 -----GCTTTCACACATPATTGATTCGATTCG 1270  
 DB 354 ThrSerLysProArgProGlnIleGlnGluValSerGlySerIleIleValIleProGln 373

QY 1269 GTTCGATGTCGGCATTCATTATAGTGTGACCTGTGGTGCAAAACAAACACCA 1210

DB 374 LysProSerArgProGlnArgProThrSerAsnProSerIleThrSerThrMetProAla 393  
 QY 1209 CAACAGTACGATGACACACG----- 1189  
 DB 394 GlnGluValIleThrThrThrThrThrThrPheLysProValIleThrSerThrGluProGln 413  
 QY 1188 -----TACAGTACAGATGAGAAAGAAACAAACACA 1159  
 DB 414 ThrThrGluProProLaseGluSerThrValIleThrThrThrSerGluProGluProThrThr 433  
 QY 1158 ACGGACAAACAAACAAACACACAGTGAATTCGAAATATTAATCACTGGACACATATCCACC 1099  
 DB 434 ThrSerSer-----ThrGlnLaseThrThrThrSerThrGlu 446  
 QY 1098 CTTTATACCCACACACAAACATAGTCCACATACATACATGTCATATATTCAGCA 1039  
 DB 447 ThrThrThrThr-----Thr 451  
 QY 1038 TCAAGTCACTCTCCATCTCACACCTCTCCAAACATATCACTCAACATGACCCCTAC 979  
 DB 452 SerThrGluAlaProThr-ThrThrValThrAlaProThrThrThrSerIleAlaProAl 471  
 QY 978 TTATGATGACACACACACACTTCATGCTTGCATAGATATTCATGAGTCTCTCCCTCC 919  
 DB 471 AlasThrThr-----AlaIleLeu 477  
 QY 918 AAGTGTCAATCATCCATCATCATCATCATCATCATCATCATCATCATCATCATCATCAT 859  
 DB 477 ProAlaLeuIleHisGln--LysLeuHisAsnLysProAlaAsnAlaPro----- 493  
 QY 858 CTCACATGACGTCACCTCAGCTCAGTCCACCAACAAAT--GACCGAGATGGAATTCACCTGTC 802  
 DB 494 -----LeuSerGlnGlnHisAsnHisGlyLeuGlyProValHisIleThrThr 510  
 QY 801 TACTGAACTGTCA-----CAAAGTCATCAACACACAGTTGTTATTTATCAAGC 754  
 DB 510 ThrGlnLaseSerLysIleValGlnGlyLysProGlnThrGluThrGlyIleValIleVal 530  
 QY 753 TCTGACCAATCAACACCGGATGAGTACGACAGACACACACAAACAGATACACAG 694  
 DB 530 ILeHisSerSerThrThr-----Th 537  
 QY 693 ACACAGAAATTAATGACATGACATCTGACGACGACGATGATGTTATTACACCA 634  
 DB 537 HisLys----- 539  
 QY 633 ACTCACTTTCTCAATATTCACCATTTCTTACGTACGAGATCAGGTGTTATCATCCT 574  
 DB 540 ---ProPheThrProLaseProSerLysAsnAspAlaValGlnAsnValSerThrLe 558  
 QY 573 T-----CCATACACACACATG-----AAAACACGATTC 543  
 DB 558 IleThrThrProProValGlnHisLysProLeuLysLysThrProLeuPro 575

RESULT 37  
 019023 PRELIMINARY; PRT; 257 AA.  
 AC 019023;  
 DT 01-JAN-1998 (TRENBLREL. 05, Created)  
 DT 01-JAN-1998 (TRENBLREL. 05, Last sequence update)  
 DT 01-MAR-2002 (TRENBLREL. 20, Last annotation update)  
 DE Elasease (EC 3.4.21.36) (Fragment).  
 GN ELa1.  
 OS Macaca mulatta (Rhesus macaque).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Cercopithecoidea;  
 OC Cercopithecinae; Macaca.  
 OX NCBI\_Taxid=9544;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=PANCREAS;  
 RA Szlegoleit A.;

RT	"Pancreatic elastase from rhesus monkey."			
RL	Submitted (JUN-1997) to the EMBL/GenBank/DBJ databases.			
CC	-1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE			
CC	TRYPsin FAMILY.			
DR	EMBL; AJ000067; CA003899.1; -.			
DR	HSSP; P05805; 1F0N.			
DR	MEROPS; S01.154; -.			
DR	InterPro; IPR001314; Chymotrypsin.			
DR	InterPro; IPR001254; Ser_protease_Try.			
DR	Pfam; PF00089; Trypsin; 1.			
DR	PRINTS; PR00722; CHYMOTRYPsin.			
DR	SMART; SM00020; TRYp_SPC; 1.			
DR	PROSITE; PS00240: TRYPsin DOM; 1.			
DR	PROSITE; PS00134; TRYPsin_HIS; UNKNOWN_1.			
DR	PROSITE; PS00135; TRYPsin_SER; 1.			
FW	Hydrolase; Serine protease.			
KT	NON_TER 1			
FT	NON_TER 1			
SO	SEQUENCE 257 AA; 27687 MW; 4D43DB67233DBDC CRC64;			
Alignment Scores:				
Pred. NO.:	0.00555	Length:	257	
Score:	114.50	Matches:	37	
Percent Similarity:	35.54%	Conservative:	22	
Best Local Similarity:	22.29%	Mismatches:	44	
Query Match:	4.26%	Indels:	63	
DB:	6	Gaps:	3	
US-10-020-441-1 (1-1477) x 019023 (1-257)				
QY	91	TGTACAGSTTCACATAGTCTCAACAGACGACTACTCAACGTGGTCAATGGTTGCTCA	150	
Db	45	CyselgylSerLeuileAlaProAspTTPvalValThrIalagIhIscySileSerSer	64	
QY	151	CCATTGCGACGATTCGGGTAAGAGATCGACCTGAAACATTTGCTCATATGAATCGA	210	
Db	65	SerLeu-----	66	
QY	211	TGTATTTCACAGACAGTGTTCGTCGTGCGTTCGTTGTGTTGTATGACATGGGTTT	270	
Db	66	-----	66	
QY	271	GCATAGATTTAATGCTCATATCTACTGTCTTAATTTACAGTTTCATT-----	318	
Db	67	-----ThyTgrInValAlaLeuGlyAspTyrAsn	76	
QY	319	CTCACACTGAGAGAAATGGCGACCAAGACATCATACCAACCGCTGGAGTTAAGGTC	378	
Db	77	LeuAlaValIlySerGluGlyProGluGlnValIleProIleAsnSerGlyAspLeuPheVal	96	
QY	379	GCACACGAGTACATGCCCTCTTGTATGTGCGACGACGAGAGAGAACCAATCGCACAGCA	438	
Db	97	HISProLeuTTPanArgLeuCysValAlaLys-----	107	
QY	439	CTCACTGATTCGATATTGCAATTGTATAGTGTGCTCAATAGTCAACTTACAGAGTGA	498	
Db	108	-----GlyAsnAspIleAlaLeuIleLysLeuSerArgSerAlaGlnLeuGlyAspAla	125	
QY	499	ATCAGAGATGATCAGTCGCGCACACGATCGGATATCCCGCACCTGGAACATGGATTTTC	558	
Db	126	ValGlnLeuAlaSerLeuProAlaGlyAspIleLeuProAsnGlnThrProCysTyr	145	
QY	559	ATTGTTGGTTATGGAAG 576		
Db	146	IleThrGlyTTPGlyArg 151		
RESULT 38				
Q9XSM1				
Q9XSM1	PRELIMINARY;	PRT;	273	AA.
AC	O9XSM1;			
DT	01-NOV-1999 (Tremblrel_12, Created)			
DT	01-NOV-1999 (Tremblrel_12, Last sequence update)			
DT	01-JUN-2002 (Tremblrel_21, Last annotation update)			

DB	SEQUENCE	Score	Length	Matches	Conservative	Mismatches	Indels	Gaps
US	10-020-441-1 (1-1477) x Q9XSM1 (1-273)	0.00633	273	114.00	34.62%	25	37	82
QY	82 AGAACAATGTGTACAGSTTCACTAGTCTCAACGAGACAGTCTACAGTGTGATTTGT	114.00	273	114.00	34.62%	25	37	82
DB	54 ArgHsgInCysgVgLySerLeuIleHsProIleThrValIleuHsIleCysE	114.00	273	114.00	34.62%	25	37	82
QY	142 GTTTGCTCAGCATTTGCCAGT-----ATTGGGTAAAGATCGACGTGAACATTT	114.00	273	114.00	34.62%	25	37	82
DB	74 IleqLyProGluLeuGInGluProSerAspPheArgValGlnLeuAdgGluGlnHsIleu	114.00	273	114.00	34.62%	25	37	82
QY	193 GTTGCTCAATGATTCATGATGATTTTCAACAGACAGTGTTCGGTGGTGTGTGT	114.00	273	114.00	34.62%	25	37	82
DB	94 TyrTyroIn--AspArgGluLeu-----	114.00	273	114.00	34.62%	25	37	82
QY	253 TGTGTATGCAAGTGGTGTTCATAGATTTTATGTCATATATCTTATTTTCAAGTT	114.00	273	114.00	34.62%	25	37	82
DB	100 -----	114.00	273	114.00	34.62%	25	37	82
QY	313 TCATTTTTCACACTGAGGAATGGCGACCAACAGGATCCATCAACCGTCTGGAGTT	114.00	273	114.00	34.62%	25	37	82
DB	100 -----	114.00	273	114.00	34.62%	25	37	82
QY	373 AAGGTGGACACAGATACATGACCCCTTTGATGTGCGACAGACAGAGACCAATGCA	114.00	273	114.00	34.62%	25	37	82
DB	101 -----	114.00	273	114.00	34.62%	25	37	82
QY	433 CAGACACTC-----AGTGAATTCGATATTTGCAATTTGA	114.00	273	114.00	34.62%	25	37	82
DB	104 ArgValIleProHsIProHsItyrTyMetValGluHsInelYalAspIleIleIleu	114.00	273	114.00	34.62%	25	37	82
QY	466 ATGTGCTGCTCAATGTGCAACTTACAGAGTGAATCAAGAGTATCACTGTGCCACAGCA	114.00	273	114.00	34.62%	25	37	82





**THIS PAGE BLANK (USPTO)**



GenCore version 5.1.4 p5 4578  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 9, 2003, 12:59:07 ; Search time 13.377 Seconds  
(without alignments)  
159.378 Million cell updates/sec

Title: US-10-020-441-3

Perfect score: 16  
Sequence: 1 VGYGRDNDNRDPSRKN 16

Scoring table: OLIGO  
Gapop 60.0 , Gapext 60.0

Searched: 908470 seqs, 133250620 residues

Word size : 0

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 1000 summaries

Database : A\_Geneseq\_101002:\*

```
1: /SIDB2/gcgdata/geneseq/geneseq-emb1/AA1980.DAT:*
2: /SIDB2/gcgdata/geneseq/geneseq-emb1/AA1981.DAT:*
3: /SIDB2/gcgdata/geneseq/geneseq-emb1/AA1982.DAT:*
4: /SIDB2/gcgdata/geneseq/geneseq-emb1/AA1983.DAT:*
5: /SIDB2/gcgdata/geneseq/geneseq-emb1/AA1984.DAT:*
6: /SIDB2/gcgdata/geneseq/geneseq-emb1/AA1985.DAT:*
7: /SIDB2/gcgdata/geneseq/geneseq-emb1/AA1986.DAT:*
8: /SIDB2/gcgdata/geneseq/geneseq-emb1/AA1987.DAT:*
9: /SIDB2/gcgdata/geneseq/geneseq-emb1/AA1988.DAT:*
10: /SIDB2/gcgdata/geneseq/geneseq-emb1/AA1989.DAT:*
11: /SIDB2/gcgdata/geneseq/geneseq-emb1/AA1990.DAT:*
12: /SIDB2/gcgdata/geneseq/geneseq-emb1/AA1991.DAT:*
13: /SIDB2/gcgdata/geneseq/geneseq-emb1/AA1992.DAT:*
14: /SIDB2/gcgdata/geneseq/geneseq-emb1/AA1993.DAT:*
15: /SIDB2/gcgdata/geneseq/geneseq-emb1/AA1994.DAT:*
16: /SIDB2/gcgdata/geneseq/geneseq-emb1/AA1995.DAT:*
17: /SIDB2/gcgdata/geneseq/geneseq-emb1/AA1996.DAT:*
18: /SIDB2/gcgdata/geneseq/geneseq-emb1/AA1997.DAT:*
19: /SIDB2/gcgdata/geneseq/geneseq-emb1/AA1998.DAT:*
20: /SIDB2/gcgdata/geneseq/geneseq-emb1/AA1999.DAT:*
21: /SIDB2/gcgdata/geneseq/geneseq-emb1/AA2000.DAT:*
22: /SIDB2/gcgdata/geneseq/geneseq-emb1/AA2001.DAT:*
23: /SIDB2/gcgdata/geneseq/geneseq-emb1/AA2002.DAT:*
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	16	100.0	16	21	AAV67588
2	16	100.0	106	21	AAV67587
3	9	56.2	11	16	AAW21484
4	6	37.5	30	22	ABB37761
5	6	37.5	30	23	ABG40652
6	6	37.5	66	22	AAU42617
7	6	37.5	95	19	AAW48217
8	6	37.5	95	19	AAW49996
9	6	37.5	95	22	AAU79051
10	6	37.5	98	22	AAU01512

11	6	37.5	98	22	AAU01513	Propeptide of cono
12	6	37.5	127	21	AAU50232	Proprionbacterium
13	6	37.5	168	21	AA829469	Cucumber xylen sap
14	6	37.5	412	22	AA866454	Protein encoded by
15	6	37.5	985	22	ABBS9214	Drosophila melanog
16	6	37.5	4215	22	ABBS9065	Drosophila melanog
17	5	31.2	15	22	AA666792	Human endoprotease
18	5	31.2	18	15	AA857838	Irb/Ilra-17, bindi
19	5	31.2	18	19	AAW75254	Fragment of human
20	5	31.2	19	22	AA663536	Amino acid sequenc
21	5	31.2	40	23	AAW50812	PS18 prostate mar
22	5	31.2	51	22	AAU39143	Proprionbacterium
23	5	31.2	51	22	AAU64366	Proprionbacterium
24	5	31.2	51	22	AAU66009	Proprionbacterium
25	5	31.2	51	22	AB827282	Protein #9281 enco
26	5	31.2	52	22	ABBA0932	Peptide #8438 enco
27	5	31.2	52	22	ABBS5053	Protein #7052 enco
28	5	31.2	52	22	AAW61794	Human brain expres
29	5	31.2	52	22	AAW74593	Human bone marrow
30	5	31.2	52	22	AAW20353	Peptide #6787 enco
31	5	31.2	52	22	AAW44705	Peptide #8742 enco
32	5	31.2	52	23	ABG44431	Human peptide enco
33	5	31.2	54	23	ABP32718	Human ORF1691 prot
34	5	31.2	57	22	AAU40965	Proprionbacterium
35	5	31.2	62	22	ABG26481	Novel human diagno
36	5	31.2	63	23	ABP05665	Human ORFX protein
37	5	31.2	68	22	ABB39553	Peptide #7059 enco
38	5	31.2	68	22	AAW60252	Human brain expres
39	5	31.2	68	22	AAW72879	Human bone marrow
40	5	31.2	68	22	AAW33110	Peptide #7147 enco
41	5	31.2	68	23	ABG42713	Human peptide enco
42	5	31.2	69	22	AAW65153	Human immune haema
43	5	31.2	72	22	ABH17825	Human nervous syst
44	5	31.2	73	22	ABBS6529	Human testicular a
45	5	31.2	73	22	AAU67927	Proprionbacterium
46	5	31.2	73	22	AAW95998	Human reproductiv
47	5	31.2	77	22	ABG14311	Novel human diagno
48	5	31.2	78	20	AAV12285	Human 5' EST seque
49	5	31.2	79	21	AAG57157	Arabidopsis thalia
50	5	31.2	79	21	AA660624	Arabidopsis thalia
51	5	31.2	80	22	AAW94793	Human reproductiv
52	5	31.2	81	22	AAU61558	Proprionbacterium
53	5	31.2	83	23	ABP09705	Human ORFX protein
54	5	31.2	84	21	AA657156	Arabidopsis thalia
55	5	31.2	84	21	AA660623	Arabidopsis thalia
56	5	31.2	84	22	ABBS6475	Human testicular a
57	5	31.2	84	22	AAW95944	Human reproductiv
58	5	31.2	84	23	ABP34730	Human ORFX703 prot
59	5	31.2	85	22	AAU58744	Proprionbacterium
60	6	31.2	88	22	AAU29639	Novel human secret
61	5	31.2	88	23	ABP03360	Human ORFX protein
62	5	31.2	96	22	AA666791	Human endoprotease
63	5	31.2	102	22	AAW59244	Human endoprotease
64	5	31.2	103	21	AAW29587	Human ion channel
65	5	31.2	103	22	ABBS9815	Amino acid sequenc
66	5	31.2	103	23	AAW67467	Human polypeptide
67	5	31.2	106	22	AAU53718	Proprionbacterium
68	5	31.2	109	23	ABP00819	Proprionbacterium
69	5	31.2	111	21	AAW24502	Human ORFX protein
70	5	31.2	112	21	AAW64155	Arabidopsis thalia
71	5	31.2	112	21	AAW60622	Arabidopsis thalia
72	5	31.2	114	22	AAU55902	Proprionbacterium
73	5	31.2	115	22	AAW69950	Human immune/haema
74	5	31.2	116	21	AAW41975	Human ORFX ORF1739
75	5	31.2	117	22	AAW06403	Human foetal prote
76	5	31.2	119	22	AAW64644	Human reproductiv
77	5	31.2	119	23	ABP08379	Human ORFX protein
78	5	31.2	123	21	AAV76001	Murine skin-cell p
79	5	31.2	123	22	AAW55940	Skin cell protein,
80	5	31.2	123	23	ABW72140	Murine protein iso
81	5	31.2	126	22	AAW75810	Human colon cancer
82	5	31.2	130	22	AAU49985	Proprionbacterium
83	5	31.2	133	17	AAW03983	DEL1-DEP2-spacer-e







522	4	25.0	11	16	AAW21497	Hepatitis delta an
523	4	25.0	11	16	AAW78499	Synthetic HIV pep
524	4	25.0	11	18	AAW28861	HIV-1a.c derived
525	4	25.0	11	20	AAW48210	Immunogenic peptid
526	4	25.0	11	21	AAW01591	Peptide comprising
527	4	25.0	11	22	AAW78241	Peptide precursor
528	4	25.0	11	23	AAW85717	Nucleophile linked
529	4	25.0	12	19	AAW20410	Human microtubule
530	4	25.0	12	20	AAW01544	TAT domain peptide
531	4	25.0	12	22	AAW59991	Internalising pept
532	4	25.0	12	23	ABP46934	Human Blys binding
533	4	25.0	13	16	AAW74799	Human tumour neuro
534	4	25.0	13	19	AAW57035	Enzyme inhibitor p
535	4	25.0	13	19	AAW56807	Enzyme inhibitor p
536	4	25.0	13	20	AAW33006	Human serine prote
537	4	25.0	13	21	AAW69964	TAT domain. Synth
538	4	25.0	13	22	AAW05269	Human immunodefici
539	4	25.0	13	22	AAW20462	Human PSK peptide
540	4	25.0	13	22	AAW66651	Antigenic peptide
541	4	25.0	13	22	AAW55699	Peptide associated
542	4	25.0	13	22	AAW55703	Peptide associated
543	4	25.0	13	22	AAW57546	HIV tat peptide fr
544	4	25.0	13	22	AAW57956	HIV tat peptide fr
545	4	25.0	13	23	AAE20724	Human M1snl extra-
546	4	25.0	13	23	AAE21025	Human Icrac-M1snl
547	4	25.0	13	23	AAO14615	Positively charged
548	4	25.0	13	23	AAO14616	Positively charged
549	4	25.0	13	23	AAW85719	Nucleophile linked
550	4	25.0	13	23	AAW69626	Cell death protect
551	4	25.0	14	19	AAW13870	Spermene modified
552	4	25.0	14	22	AAW97656	Human peptide #931
553	4	25.0	14	22	AAW00309	Human protein frag
554	4	25.0	14	22	AAW00310	Cobra venom factor
555	4	25.0	14	23	AAE21965	HIV tat protein fr
556	4	25.0	14	23	AAW47461	Residues tatCG47-
557	4	25.0	15	15	AAW49576	Rat HCNp precursor
558	4	25.0	15	15	AAW49551	Transactivating pr
559	4	25.0	15	15	AAW61311	l-Sucrose:sucrose
560	4	25.0	15	17	AAW97239	Japan cedar pollen
561	4	25.0	15	17	AAW00460	HIV-1 based transp
562	4	25.0	15	17	AAW97895	Cys-Gly-Gly linker
563	4	25.0	15	18	AAW31206	HIV tat protein fr
564	4	25.0	15	18	AAW32915	HIV Type I TAT pro
565	4	25.0	15	18	AAW26447	Tax protein fragme
566	4	25.0	15	18	AAW6152	Variant human pota
567	4	25.0	15	19	AAW41219	Human SR2 peptide
568	4	25.0	15	19	AAW41219	Human protease pep
569	4	25.0	15	20	AAW34137	Human protease pep
570	4	25.0	15	20	AAW27168	Human protease pep
571	4	25.0	15	21	AAW67144	Human protease pep
572	4	25.0	15	21	AAW67145	Human protease pep
573	4	25.0	15	21	AAW67146	Human protease pep
574	4	25.0	15	21	AAW67147	Human protease pep
575	4	25.0	15	21	AAW54781	Human subtitilisin f
576	4	25.0	15	21	AAW54782	Human subtitilisin f
577	4	25.0	15	21	AAW54783	Human subtitilisin f
578	4	25.0	15	21	AAW54784	Human subtitilisin f
579	4	25.0	15	22	AAW09933	FTTC construct. U
580	4	25.0	15	22	AAW09813	FTTC-labellled prot
581	4	25.0	15	22	AAE13065	FTTC-labellled cons
582	4	25.0	15	22	AAE12892	HIV tat protein fr
583	4	25.0	15	22	AAW65674	Human subtitilisin
584	4	25.0	15	22	AAW50222	Human subtitilisin
585	4	25.0	15	22	AAW38683	Human subtitilisin
586	4	25.0	15	22	AAW38684	Human subtitilisin
587	4	25.0	15	22	AAW38685	Human subtitilisin
588	4	25.0	15	22	AAW38686	Human subtitilisin
589	4	25.0	15	22	AAW85848	Human subtitilisin
590	4	25.0	15	22	AAW05526	Human subtitilisin
591	4	25.0	15	22	AAW04301	Human subtitilisin
592	4	25.0	15	22	AAW03419	Human subtitilisin
593	4	25.0	15	22	AAW03731	Human subtitilisin
594	4	25.0	15	22	AAW03816	Human subtitilisin





960	4	25.0	44	22	AAE04252	Human gene 4 encod
961	4	25.0	44	23	ABG46752	Human peptide enco
962	4	25.0	45	9	AAE81952	Probable C-termina
963	4	25.0	45	18	AAW25496	Random peptide rec
964	4	25.0	45	21	AAV76905	Xenopus laevis Del
965	4	25.0	45	22	ABB38447	Peptide #5953 enco
966	4	25.0	45	22	ABB39689	Peptide #7195 enco
967	4	25.0	45	22	ABB23614	Protein #5613 enco
968	4	25.0	45	22	ABB24351	Human secreted pro
969	4	25.0	45	22	ABB11306	Human brain expres
970	4	25.0	45	22	AAE59058	Human brain expres
971	4	25.0	45	22	AAE60407	Human bone marrow
972	4	25.0	45	22	AAW71591	Human bone marrow
973	4	25.0	45	22	AAW73043	Human bone marrow
974	4	25.0	45	22	AAW74065	Human bone marrow
975	4	25.0	45	22	AAW19226	Peptide #5660 enco
976	4	25.0	45	22	AAW19816	Peptide #5925 enco
977	4	25.0	45	22	AAW31888	Peptide #7302 enco
978	4	25.0	45	22	AAW33265	Peptide #11428 enc
979	4	25.0	45	22	AAW37391	Human peptide enco
980	4	25.0	45	23	ABG41403	Human peptide enco
981	4	25.0	45	23	ABG42887	Human peptide enco
982	4	25.0	45	23	ABG43961	Human peptide enco
983	4	25.0	45	23	ABG46255	Human peptide enco
984	4	25.0	46	22	ABG21117	Novel human diagno
985	4	25.0	46	22	ABB38060	Peptide #5566 enco
986	4	25.0	46	22	ABB40038	Peptide #7544 enco
987	4	25.0	46	22	ABB23283	Protein #5282 enco
988	4	25.0	46	22	AAE58695	Human brain expres
989	4	25.0	46	22	AAE60788	Human brain expres
990	4	25.0	46	22	AAW71202	Human bone marrow
991	4	25.0	46	22	AAW73462	Human immune/haema
992	4	25.0	46	22	AAW89197	Peptide #5352 enco
993	4	25.0	46	22	AAW18918	Peptide #5520 enco
994	4	25.0	46	22	AAW31483	Peptide #7697 enco
995	4	25.0	46	22	AAW33660	Human peptide enco
996	4	25.0	46	23	ABG40999	Human peptide enco
997	4	25.0	46	23	ABG43321	Human peptide enco
998	4	25.0	47	16	AAE80176	MSP1-derived poly
999	4	25.0	47	16	AAE80178	MSP1-derived poly
1000	4	25.0	47	18	AAW17971	RAC protein kinase

# ALIGNMENTS

RESULT 1					
ID	AAE67588	standard; Protein; 16 AA.			
XX	AAE67588;				
AC	19-JUN-2000	(first entry)			
XX					
DT					
XX					
DE	S. mansoni cercarial elastase protein fragment.				
XX					
KW	Schistosoma parastley; fusion protein; cercarial elastase; vaccine;				
KM	antibody response; schistosomicide.				
XX					
OS	Schistosoma mansoni.				
XX					
PN	EP992582-A2.				
XX					
PD	12-APR-2000.				
XX					
PF	05-OCT-1999;	99EP-0307832.			
XX					
PR	07-OCT-1998;	98GB-0021821.			
XX					
PA	(UYWA-) UNIV WALES BANGOR.				
XX					
PI	Doenhoff M, Sayers J;				
XX					

DR	WPI; 2000-259136/23.				
DR	N-PSDB; AA290608.				
XX					
PT	New vaccine for treatment of Schistosoma infections contains a				
PT	recombinant fusion protein comprising cercarial elastase sequence fused				
PT	to bacterial, phage or viral protein -				
XX					
PS	Claim 5; Page 16; 26pp; English.				
XX					
CC	The invention provides a vaccine comprising a recombinant fusion protein				
CC	capable of eliciting immunity against Schistosoma parasites. The fusion				
CC	protein comprises the 27 or 28 kDa cercarial elastase sequence of				
CC	S. mansoni or an active fragment, homologue or variant, fused to a				
CC	bacterial, phage or viral protein. The vaccine containing the fusion				
CC	protein has been found to induce a significantly increased antibody				
CC	response against schistosoma infections, compared to the use of S.				
CC	mansoni cercarial elastase in its native form. The present sequence				
CC	represents the S. mansoni cercarial elastase protein fragment encoded by				
CC	exon 1 of the cercarial elastase gene, used in the fusion protein of the				
CC	invention.				
XX					
XX	Sequence 16 AA;				
XX					
XX	Query Match	100.0%;	Score 16;	DB 21;	Length 16;
XX	Best Local Similarity	100.0%;	Pred. No. 1.5e-10;		
XX	Matches 16; Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;	
XX					
OY	1 VGYGRDNDPDRSKN 16				
DB	1 VGYGRDNDPDRSKN 16				

RESULT 2					
ID	AAE67587	standard; Protein; 106 AA.			
XX	AAE67587;				
AC	19-JUN-2000	(first entry)			
XX					
DT					
XX					
DE	S. mansoni cercarial elastase protein fragment.				
XX					
KW	Schistosoma parastley; fusion protein; cercarial elastase; vaccine;				
KM	antibody response; schistosomicide.				
XX					
OS	Schistosoma mansoni.				
XX					
PN	EP992582-A2.				
XX					
PD	12-APR-2000.				
XX					
PF	05-OCT-1999;	99EP-0307832.			
XX					
PR	07-OCT-1998;	98GB-0021821.			
XX					
PA	(UYWA-) UNIV WALES BANGOR.				
XX					
PI	Doenhoff M, Sayers J;				
XX					
DR	WPI; 2000-259136/23.				
XX					
PD	N-PSDB; AA290608.				
XX					
PT	New vaccine for treatment of Schistosoma infections contains a				
PT	recombinant fusion protein comprising cercarial elastase sequence fused				
PT	to bacterial, phage or viral protein -				
XX					
PS	Claim 3; Page 17; 26pp; English.				
XX					
CC	The invention provides a vaccine comprising a recombinant fusion protein				
CC	capable of eliciting immunity against Schistosoma parasites. The fusion				
CC	protein comprises the 27 or 28 kDa cercarial elastase sequence of				
CC	S. mansoni or an active fragment, homologue or variant, fused to a				
CC	bacterial, phage or viral protein. The vaccine containing the fusion				
CC	protein has been found to induce a significantly increased antibody				
CC	response against schistosoma infections, compared to the use of S.				
CC	mansoni cercarial elastase in its native form. The present sequence				
CC	represents the S. mansoni cercarial elastase protein fragment encoded by				
CC	exon 1 of the cercarial elastase gene, used in the fusion protein of the				
CC	invention.				



protein has been found to induce a significantly increased antibody response against schistosoma infections, compared to the use of S. mansoni cercarial elastase in its native form. The present sequence represents the S. mansoni cercarial elastase protein fragment encoded by exon 2 of the cercarial elastase gene, used in the fusion protein of the invention.

Sequence 106 AA;

Query Match 100.0%; Score 16; DB 21; Length 106;  
Best Local Similarity 100.0%; Pred. No. 7.8e-10;  
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VGYGRDNDRDPSRKN 16  
|||  
Db 85 VGYGRDNDRDPSRKN 100

## RESULT 3

AAW21484  
ID AAW21484 standard; peptide; 11 AA.

AC AAW21484;

DT 30-JUN-1997 (first entry)

DE Schistosoma elatase precursor derived signal oligopeptide #4.

XX Hydrophilic; signal oligopeptide; hydrophilicity maxima; vaccine; HIV;  
XX competitive inhibitor; feedback regulator; synthesis; galectin precursor;  
XX charge; polarity; farnesyl synthetase; plasminogen activator inhibitor 1;  
XX hydroxymethylglutaryl coenzyme A reductase; glucagon precursor; rhesus;  
XX gonadoliberin precursor; plasminogen activator inhibitor 2; prorenin;  
XX Alzeimer amyloid A4; corticotropin releasing factor binding protein;  
XX apolipoprotein E; herpes virus 1 glycoprotein B; HSV1; human; CMVVS;  
XX herpes virus 2 glycoprotein B; HSV2; collagenase; apolipoprotein A;  
XX triponema pallidum membrane protein; TMPA; islet amyloid polypeptide;  
XX fibroblast MPPI; schistosoma elatase precursor; schistosomn;  
XX hepatitis delta antigen; rev protein; HIV; VILV; angiotensinogen.

OS Schistosoma sp.

XX MO9519568-A1.

XX 20-JUN-1995.

PD 12-JAN-1995; 95MO-US00575.

PR 14-JAN-1994; 94US-0182248.

PA (RATH/) RATH M.

PI Rath M;

DR WPI; 1995-263953/34.

XX Identifying signal oligopeptide(s) in protein sequence(s) - shown as  
XX regions of max. hydrophilicity, used in modulating communication  
XX between protein(s)

PS Claim 5; Page 70; 88pp; English.

XX The sequences given in AAW21201-560 represent hydrophilic signal oligo-  
XX peptides. These signal oligopeptides are localised on the surface  
XX of the protein and are represented by the hydrophilicity maxima of  
XX the protein. These peptides are enriched in charged amino acids  
XX arranged with neutral spacer amino acids. The specific signal  
XX character of these oligopeptides is determined by a characteristic  
XX combination of conformation and charge within the signal sequence.  
XX These oligopeptides may be used as vaccines in the treatment of  
XX human disease, as competitive inhibitors to prevent or reduce the  
XX metabolic action or interaction of a selected protein by blocking  
XX its specific signal sequences, or as therapeutic agents to function

as feedback regulators to reduce synthesis rate of a selected protein.  
XX These peptides may be modified by omitting one or more amino acids at  
XX the N- and/or C-terminal, by substituting one or more amino acids  
XX without consideration of charge and polarity, by substituting one or  
XX more amino acids with amino acid residues with similar charge and/or  
XX polarity, by omitting one or more amino acids or a combination of these.

Sequence 11 AA;

Query Match 56.2%; Score 9; DB 16; Length 11;  
Best Local Similarity 100.0%; Pred. No. 0.003;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 5 RDDNDRDPS 13  
|||  
Db 1 RDDNDRDPS 9

## RESULT 4

ABB37761  
ID ABB37761 standard; Peptide; 30 AA.

AC ABB37761;

DT 04-FEB-2002 (first entry)

DE Peptide #5267 encoded by human foetal liver single exon probe.

XX Human; foetal liver; gene expression; single exon nucleic acid probe.

XX Homo sapiens.

XX MO200157277-A2.

XX 09-AUG-2001.

PD 30-JAN-2001; 2001WO-US00669.

PR 04-FEB-2000; 2000US-0180312.

PR 26-MAY-2000; 2000US-0207456.

PR 30-JUN-2000; 2000US-0608408.

PR 03-AUG-2000; 2000US-0632366.

PR 21-SEP-2000; 2000US-0234687.

PR 27-SEP-2000; 2000US-0236359.

PR 04-OCT-2000; 2000GB-0024263.

XX (MOLE-) MOLECULAR DYNAMICS INC.

XX Penn SG, Hanzel DK, Chen W, Rank DR;

XX WPI; 2001-483447/52.

XX Human genome-derived single exon nucleic acid probes useful for  
XX analyzing gene expression in human fetal liver -  
XX

PS Claim 27; SEQ ID NO 30396; 639pp + sequence listing; English.

XX The invention relates to a single exon nucleic acid probe for  
XX measuring human gene expression in a sample derived from human foetal  
XX liver. The single exon nucleic acid probes may be used for predicting,  
XX measuring and displaying gene expression in samples derived from human  
XX foetal liver. The present sequence is a peptide encoded by a single exon  
XX nucleic acid probe of the invention.  
XX Note: The sequence data for this patent did not form part of the  
XX printed specification, but was obtained in electronic format directly  
XX from WIPO at ftp.wipo.int/pub/published\_pct\_sequences.

SO Sequence 30 AA;

Query Match 37.5%; Score 6; DB 22; Length 30;  
Best Local Similarity 100.0%; Pred. No. 11;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

0y	8	NDRDP5	13
Db	6	NDRDP5	11
RESULT 5			
ID	ABG40652	standard; Peptide; 30 AA.	
XX	ABG40652;		
XX	19-AUG-2002	(first entry)	
DE	Human peptide encoded by genome-derived single exon probe SEQ ID 30317.		
XX	Human; single exon probe; asthma; lung cancer; COPD; ILD;		
KW	chronic obstructive pulmonary disease; interstitial lung disease;		
KW	familial idiopathic pulmonary fibrosis; neurofibromatosis;		
KW	tuberculous scleritis; Gaucher's disease; Niemann-Pick disease;		
KW	Hermansky-Pudlak syndrome; sarcoidosis; pulmonary haemostasis;		
KW	pulmonary histiocytosis; lymphangioleiomyomatosis; Karsagen syndrome;		
KW	pulmonary alveolar proteinosis; fibrocystic pulmonary dysplasia;		
KW	primary ciliary dyskinesia; pulmonary hypertension;		
KW	hyaline membrane disease.		
XX			
OS	Homo sapiens.		
XX			
PN	WO200186003-A2.		
PD			
XX	15-NOV-2001.		
PF			
XX	30-JAN-2001; 2001WO-US00665.		
PR	04-FEB-2000; 2000US-180312P.		
PR	26-MAY-2000; 2000US-207456P.		
PR	30-JUN-2000; 2000US-0608408.		
PR	03-AUG-2000; 2000US-0632366.		
PR	21-SEP-2000; 2000US-234687P.		
PR	27-SEP-2000; 2000US-236359P.		
PR	04-OCT-2000; 2000GB-0024263.		
PA	(MOLE-) MOLECULAR DYNAMICS INC.		
XX			
XX	Penn SG, Hanzel DK, Chen W, Rank DR;		
PI	WPI, 2002-114183/15.		
DR			
XX			
PT	Spatially-addressable set of single exon nucleic acid probes, used to		
XX	measure gene expression in human lung samples -		
PS	Claim 27; SEQ ID No 30317; 634pp; English.		
XX			
CC	The invention relates to a spatially-addressable set of single exon		
CC	nucleic acid probes for measuring gene expression in a sample derived		
CC	from human lung comprising single exon nucleic acid probes having one of		
CC	12614 nucleic acid sequences mentioned in the specification, or their		
CC	complements or the 12387 open reading frames derived from the 12614		
CC	probes. Also included are a microarray comprising the novel set of		
CC	probes; the novel set of probes which hybridise at high stringency to a		
CC	nucleic acid expressed in the human lung; measuring gene expression in a		
CC	sample derived from human lung, comprising (a) contacting the array with		
CC	a collection of detectably labeled nucleic acids derived from human lung		
CC	mRNA, and (b) measuring the label detectably bound to each probe of		
CC	the array; identifying exons in a eukaryotic genome, comprising		
CC	(a) algorithmically predicting at least one exon from genomic sequences		
CC	of the eukaryote; and (b) detecting specific hybridisation of detectably		
CC	labeled nucleic acids from eukaryote lung mRNA, to a single exon probe,		
CC	having a fragment identical to the predicted exon, the probe is included		
CC	in the above mentioned microarray; assigning exons to a single gene,		
CC	comprising (a) identifying exons from genomic sequence by the method		
CC	above and (b) measuring the expression of each of the exons in several		
CC	tissues and/or cell types using hybridisation to a single exon		
CC	microarrays having a probe with the exon, where a common pattern of		

CC	expression of the exons in the tissues and/or cell types indicates that the exons should be assigned to a single gene; a peptide comprising one of 12011 sequences, mentioned in the specification, or encoded by the probes/open reading frames (ORF). The probes are used for gene expression analysis, and for identifying exons in a gene, particularly using human lung derived mRNA and for the study of lung diseases such as asthma, lung cancer, chronic obstructive pulmonary disease (COPD), interstitial lung disease (ILD), familial idiopathic pulmonary fibrosis, neurofibromatosis, tuberous sclerosis, Gaucher's disease, Cc Niemann-Pick disease, Hermansky-Pudlak syndrome, sarcoidosis, pulmonary haemangioendothelioma, pulmonary histiocytosis, lymphangioleiomyomatosis, Cc pulmonary alveolar proteinosis, Karsenger syndrome, fibrocystic pulmonary dysplasia, primary ciliary dyskinesia, pulmonary hypertension and hyaline membrane disease. The present sequence is a peptide/protein encoded by a single exon probe of the invention.
CC	Note: The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from WIPO at ftp.wipo.int/pub/published_pcf_sequences.
CC	Sequence 30 AA:
CC	Query Match 37.5%; Score 6; DB 23; Length 30;
CC	Best Local Similarity 100.0%; Pred. No. 11;
CC	Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0
QY	8 NDRDPS 13 
DB	6 NDRDPS 11
RESULT 6	
AAU42617	
ID	AAU42617 standard; Protein; 66 AA.
AC	AAU42617;
XX	
DT	27-FEB-2002 (first entry)
XX	
DE	Propionibacterium acnes immunogenic protein #3513.
XX	
KM	SAPHO syndrome; synovitis; acne; pustulosis; hyperostosis; osteomyelitis; uveitis; endophthalmitis; bone; joint; central nervous system; ELISA; inflammatory lesion; acne vulgaris; enzyme linked immunosorbent assay; dermatological; osteopathic; neuroprotectant.
KW	
KX	
OS	Propionibacterium acnes.
XX	
PN	WO200181581-A2.
XX	
PD	01-NOV-2001.
XX	
PF	20-APR-2001; 2001WO-US12865.
XX	
PR	21-APR-2000; 2000US-199047P.
PR	02-JUN-2000; 2000US-208841P.
PR	07-JUL-2000; 2000US-216747P.
XX	
PA	(CORI-) CORIXA CORP.
XX	
PI	Skelley YAW, Persing DH, Mitcham JL, Wang SS, Bhatia A;
PI	L'maisonneuve J, Zhang Y, Jen S, Carter D;
XX	
DR	WPI: 2001-616774/71.
XX	
XX	N-PSDB; AAS59518.
XX	
PT	Propionibacterium acnes polypeptides and nucleic acids useful for vaccinating against and diagnosing infections, especially useful for treating acne vulgaris -
XX	
XX	Example 1; SEQ ID No 3812; 1069pp; English.
CC	Sequences AAU9105-AAU68017 represent Propionibacterium acnes immunogenic

CC polypeptides. The proteins and their associated DNA sequences are used in  
CC the treatment, prevention and diagnosis of medical conditions caused by  
CC *P. acnes*. The disorders include SAPHO syndrome (synovitis, acne,  
CC pustulosis, hypertosis and osteomyelitis), uveitis and endophthalmitis,  
CC *P. acnes* is also involved in infections of bone, joints and the central  
CC nervous system, however it is particularly involved in the inflammatory  
CC lesions associated with acne vulgaris. A method for detecting the  
CC presence or absence of *P. acnes* in a patient comprises contacting a  
CC sample with a binding agent that binds to the proteins of the invention  
CC and determining the amount of bound protein in the sample. The  
CC polypeptides may be used as antigens in the production of antibodies  
CC specific for *P. acnes* proteins. These antibodies can be used to  
CC downregulate expression and activity of *P. acnes* polypeptides and  
CC therefore treat *P. acnes* infections. The antibodies may also be used as  
CC diagnostic agents for determining *P. acnes* presence, for example, by  
CC enzyme linked immunosorbent assay (ELISA).  
CC Note: The sequence data for this patent did not form part of the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pct\_sequences.  
XX

SO Sequence 66 AA;

Query Match 37.5%; Score 6; DB 22; Length 66;  
Best Local Similarity 100.0%; Pred. No. 23;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 7 DNDRDP 12  
|||||  
Db 61 DNDRDP 66

RESULT 7  
AAW48217  
ID AAW48217 standard; Protein; 95 AA.

XX AAW48217;

XX 30-JUN-1998 (first entry)

XX Conus quercinus conantokin.

XX Conantokin; predatory cone snail; treatment; neurologic disorder;  
XX psychiatric disorder; anticonvulsant; neuroprotective; analgesic;  
XX HIV infection; ophthalmic indication; memory; learning defect;  
XX cognitive defect.

XX Conus quercinus.

XX WO9803541-A1.

XX 29-JAN-1998.

XX 21-JUL-1997; 97WO-US12618.

XX 22-JUL-1996; 96US-0684742.

XX (COGN-) COGNETIX INC.

XX (UTAH) UNIV UTAH RES FOUND.

XX Abogadie FC, Colledge C, Cruz LJ, Hillyard DR, Jimenez E;  
XX Layer RT, McCabe RT, Olivera BM, Rivier JE, Shen GS;  
XX Walker C, Zhou L;

XX WPI; 1998-120694/11.

XX N-PSDB; AAV20511.

XX New conantokin peptide(s) - useful for e.g. treating neurologic or  
XX psychiatric disorders, or the management of pain

XX Example 4; Pages 91-92; 122pp; English.

XX The present sequence is Conus quercinus conantokin,  
XX peptide derivatives of which can be used to treat neurologic and

CC psychiatric disorders, e.g. as an anticonvulsant, neuroprotective  
CC or analgesic agent. Neurologic and psychiatric disorders include  
CC epilepsy, convulsions, neurotoxic injury (associated with  
CC conditions of hypoxia, anoxia or ischaemia, which typically follow  
CC stroke, cerebrovascular accident, brain or spinal cord trauma,  
CC myocardial infarct, physical trauma, drowning, suffocation,  
CC perinatal asphyxia or hypoglycaemic events), neurodegeneration  
CC (associated with Alzheimer's disease, senile dementia, Amyotrophic  
CC lateral Sclerosis, Multiple Sclerosis, Parkinson's disease,  
CC Huntington's disease, Down's Syndrome, Korsakoff's disease,  
CC schizophrenia, AIDS dementia, multi-infarct dementia, Binswanger  
CC dementia and neuronal damage associated with uncontrolled  
CC seizures), chemical toxicity (such as addiction, and morphine,  
CC opiate, opioid and barbiturate tolerance), pain (acute, chronic,  
CC migraine), anxiety, major depression, manic-depressive illness,  
CC obsessive-compulsive disorder, schizophrenia and mood disorders  
CC (such as bipolar disorder, unipolar depression, dysthymia and  
CC seasonal affective disorder) and dysonomia (movement disorder),  
CC sleep disorder, muscle relaxation and urinary incontinence. The  
CC peptide can also be used to treat HIV infection, ophthalmic  
CC indication and memory, learning or cognitive defects.  
XX

SO Sequence 95 AA;

Query Match 37.5%; Score 6; DB 19; Length 95;  
Best Local Similarity 100.0%; Pred. No. 31;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 6 DDNRD 11  
|||||  
Db 51 DDNRD 56

RESULT 8  
AAW4996  
ID AAW4996 standard; Protein; 95 AA.

XX AAW4996;

XX 30-JUN-1998 (first entry)

XX Conus quercinus conantokin.

XX Conantokin; predatory cone snail; treatment; neurologic disorder;  
XX psychiatric disorder; anticonvulsant; neuroprotective; analgesic;  
XX HIV infection; ophthalmic indication; memory; learning defect;  
XX cognitive defect.

XX Conus quercinus.

XX WO9803189-A1.

XX 29-JAN-1998.

XX 21-JUL-1997; 97WO-US12652.

XX 06-DEC-1996; 96US-0762377.

XX 22-JUL-1996; 96US-0684750.

XX (COGN-) COGNETIX INC.

XX Layer RT, McCabe RT, Zhou L;

XX WPI; 1998-120469/11.

XX N-PSDB; AAV17142.

XX Use of conantokin peptide(s) - for treating disorders involving  
XX excessive excitation of nerve cells by excitatory amino acids or  
XX agonists of the N-methyl-D-aspartate receptor

XX Example 4; Pages 91-92; 122pp; English.

XX The present sequence is Conus quercinus conantokin,

CC peptide derivatives of which can be used to treat neurologic and  
 CC psychiatric disorders, e.g. as an anticonvulsant, neuroprotective  
 CC or analgesic agent. Neurologic and psychiatric disorders include  
 CC epilepsy, convulsions, neurotoxic injury (associated with  
 CC conditions of hypoxia, anoxia or ischaemia, which typically follow  
 CC stroke, cerebrovascular accident, brain or spinal cord trauma,  
 CC myocardial infarct, physical trauma, drowning, suffocation,  
 CC perinatal asphyxia or hypoglycaemic events), neurodegeneration  
 CC (associated with Alzheimer's disease, senile dementia, Amyotrophic  
 CC lateral Sclerosis, Multiple Sclerosis, Parkinson's disease,  
 CC Huntington's disease, Down's Syndrome, Korsakoff's disease,  
 CC schizophrenia, AIDS dementia, multi-infarct dementia, Binswanger  
 CC dementia and neuronal damage associated with uncontrolled  
 CC seizures), chemical toxicity (such as addiction, pain (acute, chronic,  
 CC opiate, opioid and barbiturate tolerance), manic-depressive illness,  
 CC (bipolar), anxiety, major depression, manic-depressive illness,  
 CC obsessive-compulsive disorder, schizophrenia and mood disorders  
 CC (such as bipolar disorder, unipolar depression, dysthymia and  
 CC seasonal affective disorder) and dysomnia (movement disorder),  
 CC sleep disorder, muscle relaxation and urinary incontinence. The  
 CC peptide can also be used to treat HIV infection, ophthalmic  
 CC indication and memory, learning or cognitive defects.

XX Sequence 95 AA;

Query Match 37.5%; Score 6; DB 19; Length 95;  
 Best Local Similarity 100.0%; Pred. No. 31;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 6 DDNDRD 11  
 |||||  
 Db 51 DDNDRD 56

#### RESULT 9

AA679051  
 ID AAG79051 standard; Protein; 95 AA.

XX AC AAG79051;

XX DT 10-DEC-2001 (first entry)

DE Amino acid sequence of conantokin Qu precursor protein.

XX Conantokin; cone snail; nerve cell excitation; NMDA receptor; epilepsy;  
 KW N-methyl-D-aspartate receptor; pain; psychiatric disorder;  
 KW neurotoxic injury; hypoxia; anoxia; ischemia; neurodegeneration;  
 KW chemical toxicity; addiction; drug craving; psychiatric disorder;  
 KW anxiety; depression; obsessive compulsive disorder; schizophrenia;  
 KW mood disorder; ophthalmic disorder; neurological disorder; dystonia;  
 KW sleep disorder; muscle relaxation; urinary incontinence;  
 KW cognition enhancement; HIV infection.

XX Conus quercinus.

OS US6277825-B1.

XX PD 21-AUG-2001.

XX PF 20-JUL-1999; 990US-0357141.

XX PR 22-JUL-1996; 96US-0684750.

PR 06-DEC-1996; 96US-0762377.

PR 21-JUL-1997; 97WO-US12652.

PR 10-FEB-1999; 990US-0142076.

PR 01-APR-1999; 990US-0283277.

XX (UTAH ) UNIV UTAH RES FOUND.

PA (COGN-) COGNETIX INC.

XX PI Oliveira BM, McIntosh JM, McCabe RT, Layer RT, Zhou L;

XX WPI, 2001-601377/68.

DR N-PSDB; AA165379.

XX Use of conantokin peptide or its derivatives or a conantokin peptide

PT chimera for treating disorders e.g. migraine -

XX Example 4; Column 75-76; 60pp; English.

XX The present sequence represents conantokin precursor protein. Conantokins  
 CC differ from conotoxins, in that they contain gamma-carboxyglutamic acid.  
 CC The conantokins are derived from the venom of cone snails. They are used  
 CC for the treatment of disorders in which the pathophysiology involves  
 CC excessive excitation of nerve cells by excitatory amino acids or agonist  
 CC of N-methyl-D-aspartate (NMDA) receptor. The conantokin peptides are  
 CC used for the treatment of disorders such as pain; neurologic or  
 CC psychiatric disorders such as epilepsy; for reducing neurotoxic injury  
 CC associated with conditions of hypoxia, anoxia or ischemia; for treating  
 CC neurodegeneration; for treating chemical toxicity such as addiction,  
 CC drug craving, alcohol abuse, morphine, opiate and barbiturate tolerance;  
 CC for treating psychiatric disorders such as anxiety, major depression,  
 CC manic-depression illness, obsessive compulsive disorder, schizophrenia  
 CC or mood disorder; for treating ophthalmic disorder; for treating  
 CC additional neurological disorders e.g. dystonia, sleep disorder, muscle  
 CC relaxation and urinary incontinence; for memory/cognition enhancement;  
 CC for treating HIV infection.

XX Sequence 95 AA;

Query Match 37.5%; Score 6; DB 22; Length 95;  
 Best Local Similarity 100.0%; Pred. No. 31;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 6 DDNDRD 11  
 |||||  
 Db 51 DDNDRD 56

#### RESULT 10

AAU01512  
 ID AAU01512 standard; Protein; 98 AA.

XX AC AAU01512;

XX DT 29-AUG-2001 (first entry)

DE Propeptide of conopeptide Lvl, amino acid sequence.

XX Gamma carboxyglutamate; neurological disorder; epilepsy; trauma;  
 KW hypoxia; anoxia; ischaemia; stroke; brain; spinal cord; suffocation;  
 KW myocardial infarct; drowning; perinatal asphyxia; hypoglycaemia;  
 KW neurodegeneration; Alzheimer's disease; Huntington's disease;  
 KW senile dementia; Amyotrophic lateral Sclerosis; multiple sclerosis;  
 KW Parkinson's disease; Down's Syndrome; Korsakoff's disease; schizophrenia;  
 KW AIDS; acquired immunodeficiency syndrome; HIV; neuronal damage; pain;  
 KW seizure; chemical toxicity; addiction; dysomnia; psychiatric disorder;  
 KW mood disorder; memory; ophthalmic; parasitic worm; conopeptide Lvl.

XX Conus lividus.

OS Key Location/Qualifiers

FT Misc-difference 73 /note= "Encoded by ncg"

XX WO200118033-A1.

XX PD 15-MAR-2001.

XX PF 08-SEP-2000; 2000WO-US24816.

XX PR 10-SEP-1999; 99US-0153034.

XX PR 21-JUL-2000; 2000US-0219673.

XX (UTAH ) UNIV UTAH RES FOUND.

PA (COGN-) COGNETIX INC.

XX Olivera BM, McIntosh JM, Garrett JE, Walker CS, Watkins M;  
 PI Jones RM;  
 XX WPI: 2001-273379/28.  
 DR N-PSDB; AAS02199.  
 XX  
 PT New isolated gamma-carboxyglutamine containing peptide for treating or  
 PT preventing neurological and psychiatric disorders e.g. epilepsy,  
 PT Alzheimer's disease, migraine, chemical toxicity, dystonia, anxiety,  
 PT and depression -  
 XX  
 PS Claim 5; Page 35; 102pp; English.  
 XX  
 CC The sequence represents the amino acid sequence of the propeptide of  
 CC gamma carboxyglutamate-containing conopeptide Lv1. The conopeptide  
 CC is used for treating or preventing disorders in which the pathophysiology  
 CC involves excess excitation of nerve cells by excitatory amino acids or  
 CC agonists of heterogeneous ionotropic glutamate receptors or heterogeneous  
 CC G protein coupled glutamate receptors. The disorders may be neurological  
 CC disorders, such as: (i) seizure associated with epilepsy; (ii) a  
 CC neurotoxic injury associated with hypoxia, anoxia, ischaemia, stroke,  
 CC cerebrovascular accident, brain or spinal cord trauma, myocardial  
 CC infarct, physical trauma, drownings, suffocation, perinatal asphyxia, or  
 CC hypoglycaemic events; (iii) neurodegeneration associated with Alzheimer's  
 CC disease, Huntington's disease, senile dementia, Amyotrophic Lateral  
 CC Sclerosis, multiple sclerosis, Parkinson's disease, Down's Syndrome,  
 CC Korsakoff's disease, schizophrenia, AIDS (acquired immunodeficiency  
 CC syndrome) dementia from HIV infection, HIV infection, multi-infarct  
 CC dementia, Binswanger dementia and neuronal damage associated with  
 CC uncontrolled seizures; (iv) pain which is a migraine, acute pain, or  
 CC persistent pain; (v) chemical toxicity which is addiction, morphine,  
 CC opiate, opioid and barbiturate tolerance; and (vi) dystonia, urinary  
 CC incontinence, muscle relaxation or sleep disorder. The disorders may be  
 CC psychiatric disorders, such as, anxiety, major depression, manic-  
 CC depressive illness, obsessive-compulsive disorder, schizophrenia, or mood  
 CC disorders (bipolar disorder, unipolar depression, dysphymia, or seasonal  
 CC effective disorder). The conopeptide is also used to treat memory or  
 CC cognitive deficits, ophthalmic indications, or to control nematodes or  
 CC parasitic worms.  
 CC  
 XX  
 SQ Sequence 98 AA;  
 Query Match 37.5%; Score 6; DB 22; Length 98;  
 Best Local Similarity 100.0%; Pred. No. 32;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 6 DDNDRD 11  
 Db 53 DDNDRD 58  
 XX  
 XX  
 AC AAU01513;  
 XX  
 DT 29-AUG-2001 (first entry)  
 XX  
 DE Propeptide of conopeptide Lv2, amino acid sequence.  
 XX  
 KW Gamma carboxyglutamate, neurological disorder; epilepsy; trauma;  
 KW hypoxia; anoxia; ischaemia; stroke; brain; spinal cord; suffocation;  
 KW myocardial infarct; drowning; perinatal asphyxia; hypoglycaemia;  
 KW neurodegeneration; Alzheimer's disease; Huntington's disease;  
 KW senile dementia; Amyotrophic Lateral Sclerosis; multiple sclerosis;  
 KW Parkinson's disease; Down's Syndrome; Korsakoff's disease; schizophrenia;  
 KW AIDS; acquired immunodeficiency syndrome; HIV; neuronal damage; pain;  
 KW seizure; chemical toxicity; addiction; dystonia; psychiatric disorder;  
 KW mood disorder; memory; ophthalmic; parasitic worm; conopeptide Lv2.  
 OS Conus lividus.

XX  
 FH Key Location/Qualifiers  
 FT Misc-difference 27  
 FT /note= "Encoded by ant"  
 XX  
 XX WO200118033-A1.  
 XX  
 PD 15-MAR-2001.  
 XX  
 XX 08-SEP-2000; 2000WO-US24816.  
 XX  
 PR 10-SEP-1999; 99US-0153034.  
 XX  
 PR 21-JUL-2000; 2000US-0219673.  
 XX  
 PA (UTAH ) UNIV UTAH RES FOUND.  
 PA (COGN-) COGNETIX INC.  
 XX  
 PI Olivera BM, McIntosh JM, Garrett JE, Walker CS, Watkins M;  
 PI Jones RM;  
 XX WPI: 2001-273379/28.  
 DR N-PSDB; AAS02200.  
 XX  
 PT New isolated gamma-carboxyglutamine containing peptide for treating or  
 PT preventing neurological and psychiatric disorders e.g. epilepsy,  
 PT Alzheimer's disease, migraine, chemical toxicity, dystonia, anxiety,  
 PT and depression -  
 XX  
 PS Claim 5; Page 35; 102pp; English.  
 XX  
 CC The sequence represents the amino acid sequence of the propeptide of  
 CC gamma carboxyglutamate-containing conopeptide Lv2. The conopeptide  
 CC is used for treating or preventing disorders in which the pathophysiology  
 CC involves excess excitation of nerve cells by excitatory amino acids or  
 CC agonists of heterogeneous ionotropic glutamate receptors or heterogeneous  
 CC G protein coupled glutamate receptors. The disorders may be neurological  
 CC disorders, such as: (i) seizure associated with epilepsy; (ii) a  
 CC neurotoxic injury associated with hypoxia, anoxia, ischaemia, stroke,  
 CC cerebrovascular accident, brain or spinal cord trauma, myocardial  
 CC infarct, physical trauma, drownings, suffocation, perinatal asphyxia, or  
 CC hypoglycaemic events; (iii) neurodegeneration associated with Alzheimer's  
 CC disease, Huntington's disease, senile dementia, Amyotrophic Lateral  
 CC Sclerosis, multiple sclerosis, Parkinson's disease, Down's Syndrome,  
 CC Korsakoff's disease, schizophrenia, AIDS (acquired immunodeficiency  
 CC syndrome) dementia from HIV infection, HIV infection, multi-infarct  
 CC dementia, Binswanger dementia and neuronal damage associated with  
 CC uncontrolled seizures; (iv) pain which is a migraine, acute pain, or  
 CC persistent pain; (v) chemical toxicity which is addiction, morphine,  
 CC opiate, opioid and barbiturate tolerance; and (vi) dystonia, urinary  
 CC incontinence, muscle relaxation or sleep disorder. The disorders may be  
 CC psychiatric disorders, such as, anxiety, major depression, manic-  
 CC depressive illness, obsessive-compulsive disorder, schizophrenia, or mood  
 CC disorders (bipolar disorder, unipolar depression, dysphymia, or seasonal  
 CC effective disorder). The conopeptide is also used to treat memory or  
 CC cognitive deficits, ophthalmic indications, or to control nematodes or  
 CC parasitic worms.  
 CC  
 XX  
 SQ Sequence 98 AA;  
 Query Match 37.5%; Score 6; DB 22; Length 98;  
 Best Local Similarity 100.0%; Pred. No. 32;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 6 DDNDRD 11  
 Db 53 DDNDRD 58  
 XX  
 XX  
 AC AAU50232;  
 XX  
 DT 29-AUG-2001 (first entry)  
 XX  
 DE Propeptide of conopeptide Lv2, amino acid sequence.  
 XX  
 KW Gamma carboxyglutamate, neurological disorder; epilepsy; trauma;  
 KW hypoxia; anoxia; ischaemia; stroke; brain; spinal cord; suffocation;  
 KW myocardial infarct; drowning; perinatal asphyxia; hypoglycaemia;  
 KW neurodegeneration; Alzheimer's disease; Huntington's disease;  
 KW senile dementia; Amyotrophic Lateral Sclerosis; multiple sclerosis;  
 KW Parkinson's disease; Down's Syndrome; Korsakoff's disease; schizophrenia;  
 KW AIDS; acquired immunodeficiency syndrome; HIV; neuronal damage; pain;  
 KW seizure; chemical toxicity; addiction; dystonia; psychiatric disorder;  
 KW mood disorder; memory; ophthalmic; parasitic worm; conopeptide Lv2.  
 OS Conus lividus.

```

XX 13-FEB-2002 (first entry)
DT
XX
DE Propionibacterium acnes immunogenic protein #11128.
XX
KM SAPHO syndrome; synovitis; acne; pustulosis; hypertosis; osteomyelitis;
KM uveitis; endophthalmitis; bone; joint; central nervous system; ELISA;
KM inflammatory lesion; acne vulgaris; enzyme linked immunosorbent assay;
KM dermatological; osteopathic; neuroprotectant.
XX
OS Propionibacterium acnes.
XX
EN WO200181581-A2.
XX
PD 01-NOV-2001.
XX
PF 20-APR-2001; 2001WO-US12865.
XX
PR 21-APR-2000; 2000US-199047P.
PR 02-JUN-2000; 2000US-208841P.
PR 07-JUL-2000; 2000US-216747P.
XX
PA (CORI-) CORIXA CORP.
XX
PI Skelky YAM, Persing DH, Mitcham JU, Wang SS, Bhatia A;
PI L'maisonneuve J, Zhang Y, Jen S, Carter D;
XX
DR MPI: 2001-616774/71.
DR N-PSDB; AAS59548.
XX
PT Propionibacterium acnes polypeptides and nucleic acids useful for
PT vaccinating against and diagnosing infections, especially useful for
PT treating acne vulgaris -
XX
PS Example 1; SEQ ID NO 11427; 10699P; English.
XX
CC Sequences AAU93105-AAU68017 represent Propionibacterium acnes immunogenic
CC polypeptides. The proteins and their associated DNA sequences are used in
CC the treatment, prevention and diagnosis of medical conditions caused by
CC P. acnes. The disorders include SAPHO syndrome (synovitis, acne,
CC pustulosis, hypertosis and osteomyelitis), uveitis and endophthalmitis.
CC P. acnes is also involved in infections of bone, joints and the central
CC nervous system, however it is particularly involved in the inflammatory
CC lesions associated with acne vulgaris. A method for detecting the
CC presence or absence of P. acnes in a patient comprises contacting a
CC sample with a binding agent that binds to the proteins of the invention
CC and determining the amount of bound protein in the sample. The
CC polypeptides may be used as antigens in the production of antibodies
CC specific for P. acnes proteins. These antibodies can be used to
CC downregulate expression and activity of P. acnes polypeptides and
CC therefore treat P. acnes infections. The antibodies may also be used as
CC diagnostic agents for determining P. acnes presence, for example, by
CC enzyme linked immunosorbent assay (ELISA).
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 127 AA:

```

```

Query Match 37.5%; Score 6; DB 22; Length 127;
Best Local Similarity 100.0%; Pred. No. 40;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 6 DDNDRD 11
DB 40 DDNDRD 45

```

```

RESULT 13
AAB29469
ID AAB29469 standard; Protein; 168 AA.
XX
AC AAB29469;

```

```

XX 09-FEB-2001 (first entry)
DT
XX
DE Cucumber xylem sap protein XSP4.
XX
KM Cucumber XSP4; xylem sap protein; intercellular fluid;
KM secretion; recombinant protein production; transgenic plant.
XX
OS Cucumis sativus.
XX
EN JP2000245470-A.
XX
PD 12-SEP-2000.
XX
PF 26-FEB-1999; 99JP-0051904.
XX
PR 26-FEB-1999; 99JP-0051904.
XX
PA (NODA ) 2H NODA SANGYO KAGAKU KENKYUSHO.
XX
DR MPI; 2000-675182/66.
DR N-PSDB; AAC63928.
XX
PT Novel method for the preparation of an exotic polypeptide by
PT transforming a plant to express and secrete the peptide -
XX
PS Example 2; Page 20-21; 33pp; Japanese.
XX
CC The invention relates to the recombinant expression of a foreign protein
CC in a transgenic plant whereby the foreign protein is secreted into the
CC intercellular fluid. The invention provides the cucumber xylem sap
CC protein XSP30 and XSP4 promoters (AAC63913, AAC63914) and a cucumber
CC XSP30 secretion signal peptide (AAB29466) for expression of the foreign
CC protein according to the method of the invention. The invention also
CC relates to expression constructs comprising a promoter of the invention
CC and DNA encoding the signal peptide, and transgenic plants and plant
CC cells comprising the expression vector. The method of the invention is
CC useful for the efficient production of a foreign protein in a plant, and
CC enables the recombinantly produced protein to be easily collected. The
CC present sequence represents cucumber XSP4.
XX
SQ Sequence 168 AA:

```

```

Query Match 37.5%; Score 6; DB 21; Length 168;
Best Local Similarity 100.0%; Pred. No. 52;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 1 VGYGRD 6
DB 53 VGYGRD 58

```

```

RESULT 14
AAB6454
ID AAB6454 standard; Protein; 412 AA.
XX
AC AAB6454;
XX
DT 09-APR-2001 (first entry)
XX
DE Protein encoded by Mycobacterium tuberculosis Rv0204c gene.
XX
KM Mycobacterium tuberculosis; attenuated microorganism; Rv0204c;
KM signature tagged transposon mutant; mutant library;
KM mycobacterial infection; actinomycetales; antibacterial; immunostimulant;
KM vaccine.
XX
OS Mycobacterium tuberculosis.
XX
EN WO200102555-A1.
XX
PD 11-JAN-2001.
XX

```

```

XX 06-JUL-2000; 2000OWO-1B00950.
PF
FR 06-JUL-1999; 99US-0142982.
FR 08-JUL-1999; 99US-0142833.
XX
XX (INSP ) INST PASTEUR.
PA
PI Gicquel B, Guilhot C, Camacho L;
XX
XX WPI; 2001-091804/10.
DR
DR N-PSDB; AAF31613.
XX
PT Screening a mutant library for mutants unable to grow under specific
PT conditions and for identifying loci involved in pathogenicity,
FT comprised using signature tagged transposon mutagenesis
XX
XX
PS Example 8; Fig 16A; 159pp; English.
XX
XX The present sequence is given in a specification relating to a method for
CC screening a library of mutants. The method comprises constructing a
CC library with insertions in genes and/or regulatory regions of the
CC organisms of interest, where the insertion contains a tag and/or a
CC transposon associated with a tag. The mutants are identified by
CC hybridisation of the tags to known sequences. The method is useful for
CC treating an individual suffering from a mycobacterial infection,
CC suspected of being infected with a Mycobacterium, or having been
CC exposed to an infectious Mycobacterium. It is also useful for
CC identifying and isolating mutants of actinomycetales and for identifying
CC compounds that have antibiotic activity. The method is used to identify
CC mutants of microorganisms, preferably an actinomycetales, such as
CC M. tuberculosis, M. bovis, M. leprae, M. avium, M. intracellulare and
CC M. paratuberculosis, that is unable to grow under specific conditions.
CC It is especially useful for identifying loci involved in pathogenicity.
CC It is useful in constructing vaccines. The method can be used to screen
CC multiple libraries concurrently. It can screen libraries of different
CC organisms or different strains of the same organism. The present
CC protein is encoded by a gene which is disrupted by the insertion of
CC the IS 1096 transposon to produce an attenuated mutant of Mycobacterium
CC tuberculosis.
CC
SO Sequence 412 AA;

Query Match 37.5%; Score 6; DB 22; Length 412;
Best Local Similarity 100.0%; Pred. No. 1.1e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0.

OY 7 DNDRDP 12
|||||
DB 363 DNDRDP 368

RESULT 15
ABBS9214
ID ABBS9214 standard; Protein; 965 AA.
XX
XX ABBS9214;
XX
XX 26-MAR-2002 (first entry)
XX
XX Drosophila melanogaster polypeptide SEQ ID NO 4434.
XX
XX Drosophila; developmental biology; cell signalling; insecticide;
XX pharmaceutical.
XX
XX Drosophila melanogaster.
XX
XX WO200171042-A2.
XX
XX 27-SEP-2001.
XX
XX 23-MAR-2001; 2001WO-US09231.
XX
XX 23-MAR-2000; 2000US-191637P.
XX
XX

```

```

PR      11-JUL-2000; 2000US-0614150.
XX
XX      (PEKE ) PE CORP NY.
XX
XX      Venter JC, Adams M, Li PWD, Myers EW;
XX
XX      WPI; 2001-656860/75.
XX      N-PsDB; ABL0317.
XX
XX      New isolated nucleic acid detection reagent for detecting 1000 or more
XX      genes from Drosophila and for elucidating cell signalling and cell-cell
XX      interactions -
XX
XX      Disclosure; SEQ ID NO 4434; 21np + Sequence listing; English.
XX
XX      The invention relates to an isolated nucleic acid detection reagent
XX      capable of detecting 1000 or more genes from Drosophila. The invention is
XX      useful in developmental biology and in elucidating cell signalling and
XX      cell-cell interactions in higher eukaryotes for the development of
XX      insecticides, therapeutic and pharmaceutical drugs. The invention
XX      discloses genomic DNA sequences (AB101840-AB161615), expressed DNA
XX      sequences (AB101840-AB161615) and the encoded proteins
XX      (ABB57737-ABB72072).
XX      The sequence described for this patent did not form part of the printed
XX      specification, but was obtained in electronic format directly from WIPO
XX      at ftp.wipo.int/pub/published_pat_sequences.
XX
XX      Sequence   985 AA;
SQ
Query Match          37.5%; Score 6; DB 22; Length 985;
Best Local Similarity 100.0%; Pred. No. 2.4e+02;
Matches    6; Conservative    0; Mismatches    0; Indels    0; Gaps    0
OY      1 VGYGRD 6
        |||||
Db       819 VGYGRD 824

RESULT 16
ID      ABB59065
XX      ABB59065 standard; Protein; 4215 AA.
XX
XX      ABB59065;
AC
XX
XX      26-MAR-2002 (first entry)
DT
XX
XX      Drosophila melanogaster polypeptide SEQ ID NO 3987.
DE
XX
XX      Drosophila; developmental biology; cell signalling; insecticide;
XX      pharmaceutical.
KW
XX
XX      Drosophila melanogaster.
OS
XX
XX      WO200171042-A2.
PN
XX
XX      27-SEP-2001.
PD
XX
XX      23-MAR-2001; 2001WO-US09231.
PF
XX
XX      23-MAR-2000; 2000US-191637P.
PR      11-JUL-2000; 2000US-0614150.
XX
XX      (PEKE ) PE CORP NY.
XX
XX      Venter JC, Adams M, Li PWD, Myers EW;
XX
XX      WPI; 2001-656860/75.
XX      N-PsDB; ABL03168.
XX
XX      New isolated nucleic acid detection reagent for detecting 1000 or more
XX      genes from Drosophila and for elucidating cell signalling and cell-cell
XX      interactions -
XX

```

PS Disclosure; SEQ ID NO 3987; 21bp + Sequence Listing; English.  
XX  
CC The invention relates to an isolated nucleic acid detection reagent  
CC capable of detecting 1000 or more genes from *Drosophila*. The invention is  
CC useful in developmental biology and in elucidating cell signalling and  
CC cell-cell interactions in higher eukaryotes for the development of  
CC insecticides, therapeutics and pharmaceutical drugs. The invention  
CC discloses genomic DNA sequences (AB16176-AB16176), expressed DNA  
CC sequences (AB101840-AB16175) and the encoded proteins  
CC (AB857737-AB872072).  
CC The sequence data for this patent did not form part of the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pct\_sequences.  
XX  
SQ Sequence 4215 AA;  
  
Query Match 37.5%; Score 6; DB 22; Length 4215;  
Best Local Similarity 100.0%; Pred. No. 8.7e+02;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
QY 9 DRDPS 14  
Db 3697 DRDPSR 3702  
|||||  
AAAG66792 standard; peptide; 15 AA.  
ID AAAG66792 standard; peptide; 15 AA.  
AC AAAG66792;  
DT 15-DEC-2001 (first entry)  
DE Human endoprotease 11 N-terminal peptide.  
XX  
XX Human; endoprotease 11; recombinant production;  
KM malignant tumour; cancer; blood disease; HIV infection;  
KM human immunodeficiency virus; immune disorder; inflammatory condition;  
KM cyclostatic; anti-HIV; anti-inflammatory; immunomodulator;  
KM N-terminal peptide; enzyme linked immunosorbent assay; ELISA.  
XX  
OS Homo sapiens.  
XX WO200174883-A1.  
XX  
XX 11-OCT-2001.  
XX 26-MAR-2001; 2001WO-CN00485.  
XX 28-MAR-2000; 2000CN-0115211.  
XX (SHAN-) SHANGHAI BIOWINDOW GENE DEV INC.  
XX PA  
XX Mao Y, Xie Y;  
XX WPI; 2001-639350/73.  
XX  
XX New human endoprotease 11 for diagnosing and treating cancer.  
PT hemopathy, human immunodeficiency virus infection, immunological  
PT diseases and inflammation -  
XX  
XX Example 5; Page 18; 36pp; Chinese.  
XX  
XX The invention relates to human endoprotease 11 (AA66791), nucleic acids  
CC encoding it (AAH76947), and a method for the recombinant production of  
CC endoprotease 11. The protein has a molecular weight of 11 kD. The present  
CC invention additionally discloses an antagonist of endoprotease 11  
CC for therapeutic use, and an antibody which specifically binds to  
CC endoprotease 11. Endoprotease 11, and nucleotides which encode it may  
CC be used for treating a variety of diseases, such as malignant tumours,  
CC blood diseases, HIV (human immunodeficiency virus) infection, immune  
CC disorders and inflammatory conditions. The protein may also be used to  
CC screen for modulators of its activity or for peptide fingerprinting

CC identification. The polynucleotide can be used as a primer for nucleic  
CC acid amplification reactions or as a probe for hybridisation reactions,  
CC or in producing gene chips or microarrays. The present sequence  
CC represents the 15 N-terminal amino acids of human endoprotease 11 used in  
CC ELISA (enzyme linked immunosorbent assay) in an exemplification of the  
CC invention.  
XX  
SQ Sequence 15 AA;  
  
Query Match 31.2%; Score 5; DB 22; Length 15;  
Best Local Similarity 100.0%; Pred. No. 72;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
QY 9 DRDPS 13  
Db 4 DRDPS 8  
|||||  
AAR57838 standard; peptide; 18 AA.  
ID AAR57838 standard; peptide; 18 AA.  
AC AAR57838;  
DT 28-MAR-1995 (first entry)  
DE IIB/IIIA-17, binding site for platelet glycoprotein GPIIb/IIIa.  
XX  
XX Binding site; CDR; complementarity determining region; immunoglobulin;  
KM heavy; light; primer extension; PCR; amplify; fibronectin; vitronectin;  
KM RGD-dependant; integrin ligand; von Willebrand factor; EBV; gp350/220;  
KM envelope glycoprotein; HIV; gp120; reovirus; hemagglutinin; insulin;  
KM cellular receptor; CR2; CD4; hormone; thyroid stimulating hormone; TSH;  
KM transferrin; apolipoprotein; apo E; apo A1; MHC; class I; class II;  
KM non-RGD-dependant; vitronectin receptor; alpha-V, beta-3; modulation;  
KM anti-GPIIb/IIIa; monoclonal antibody; Mab; platelet adhesion; cancer;  
KM coagulation; inflammation; anti-vitronectin; tumour cell adhesion;  
KM migration.  
XX  
OS Homo sapiens.  
XX FH  
XX Key Location/Qualifiers  
FT Misc-difference 10 /note= "Undefined amino acid"  
FT  
XX  
XX WO9418221-A.  
XX PD 18-AUG-1994.  
XX PF 02-FEB-1994; 94WO-US01258.  
XX PR 02-FEB-1993; 93US-0012566.  
XX PR 28-JUN-1993; 93US-0084542.  
XX PA (SCRI) SCRIPPS RES INST.  
XX PI Barbas CF, Lerner RA;  
XX WPI; 1994-279675/34.  
XX  
XX Production of binding sites within CDR regions of immunoglobulins  
PT - displayed on the surface of filamentous phage particles, for  
PT inhibiting platelet aggregation and vitronectin binding  
XX  
XX Claim 44; Page 22; 207pp; English.  
XX  
XX The sequences given in AAR57837-84 are binding sites which were used in  
CC the method of the invention for producing a polypeptide having a  
CC binding site capable of binding a preselected agent. Nucleotide  
CC sequences encoding these binding site peptides were introduced into  
CC a CDR region of a nucleic acid encoding an immunoglobulin heavy (H)  
CC or light (L) chain, by amplifying the CDR region by primer extension.  
CC Preferred binding sites are derived from the RGD-dependant integrin



CC ligands, eg. fibronectin, vitronectin, von Willebrand factor, from  
CC the envelope glycoprotein from viruses such as HIV gp120, EBV gp350/  
CC 220, reovirus haemagglutinin, from cellular receptors such as CR2 or  
CC CD4, from protein hormones such as thyroid stimulating hormone (TSH),  
CC insulin, transferrin, from apolipoproteins such as apo E and apo A1,  
CC from immunoglobulin GDRs and from MHC class I or II proteins. Non-RGD-  
CC dependent integrin binding sites were selected for the affinity to bind  
CC vitronectin receptor alpha-v, beta-3. An anti-gp11b/IIa monoclonal  
CC antibody (Mab) produced in this way can be used to modulate platelet  
CC adhesion in the treatment of coagulation and some inflammatory  
CC responses. An anti-vitronectin Mab can be used in the treatment of cancer  
CC by blocking tumour cell adhesion and migration. This sequence  
CC represents an RGD-dependant binding site which has been shown to bind the  
CC human platelet glycoprotein gp11b/IIa (alpha11b, beta3) when present in  
CC a phagemid display protein.  
XX  
SQ Sequence 18 AA;  
Query Match 31.2%; Score 5; DB 15; Length 18;  
Best Local Similarity 100.0%; Pred. No. 84;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 VGXGR 5  
Db 1 VGXGR 5  
RESULT 19  
AAW75254  
ID AAW75254 standard; Protein; 18 AA.  
XX  
AC AAW75254;  
XX  
DT 29-JAN-1999 (first entry)  
XX  
DE Fragment of human secreted protein encoded by gene 18.  
XX  
XX Human; secreted protein; fusion protein; gene therapy; protein therapy;  
KW diagnosis; tissue; cancer; tumour; neurodegenerative disorder; leukaemia;  
KW developmental abnormality; foetal deficiency; blood; allergy; renal;  
KW immune system; aetna; lymphocytic disease; brain; hepatic; lymphoma;  
KW inflammation; ischaemic shock; Alzheimer's disease; restenosis; AIDS;  
KW congenital disorder; schizophrenia; prostate; obesity; osteoclast; thymus;  
KW osteoporosis; arthritis; testis; lung; thyroiditis; thyroid; digestion;  
KW endocrine; metabolism; regulation; malabsorption; gastritis; neoplasm.  
XX  
OS Homo sapiens.  
XX  
XX MO9840483-A2.  
XX  
XX 17-SEP-1998.  
PD  
XX 12-MAR-1998; 98WO-US04858.  
PF  
XX 19-DEC-1997; 97US-0068368.  
PR 14-MAR-1997; 97US-0040710.  
PR 14-MAR-1997; 97US-0040762.  
PR 30-MAY-1997; 97US-0048108.  
PR 30-MAY-1997; 97US-0048189.  
PR 30-MAY-1997; 97US-0048357.  
PR 30-MAY-1997; 97US-0050934.  
PR 06-JUN-1997; 97US-0048970.  
PR 05-SEP-1997; 97US-0057765.  
XX  
XX (HUMA-) HUMAN GENOME SCI INC.  
XX  
PI Ferric AM, Fischer CL, Gentz RL, Greene JM, Kyaw H;  
PI Li H, Li Y, Moore PA, Rosen CA, Ruben SM, Soppet DR;  
PI Wei YF, Young PE, Zeng Z;  
XX  
XX WPI; 1998-520811/44.  
XX  
XX N-PSDB; AAV34303.

PT Isolated human poly(nucleotide(s) encoding secretory peptide(s) -  
PT used to develop products for the diagnosis and treatment of e.g.  
PT inflammation, cancers, CNS disorders or immune system disorders  
XX  
XX Disclosure; Page 20; 201pp; English.  
XX  
XX This sequence represents a fragment of a secreted human protein encoded  
CC by the nucleic acid molecule designated Gene 18 (AAV34303). The gene  
CC can be used to generate fusion proteins by linking to the gene to a  
CC human immunoglobulin Fc portion (e.g. AAV34277) for increasing the  
CC stability of the fused protein as compared to the human protein only.  
CC The invention relates to 28 novel genes and their fragments (nucleic  
CC acid sequences: AAV34286-V34325; amino acid sequences AAW75196-W75235)  
CC which are useful for preventing, treating or ameliorating medical  
CC conditions e.g. by protein or gene therapy. Also, pathological  
CC conditions can be diagnosed by determining the amount of the new  
CC polypeptides in a sample or by determining the presence of mutations in  
CC the new polynucleotides. Specific uses are described for each of the 28  
CC polynucleotides, based on which tissues they are most highly expressed in  
CC (see AAV34286 for described uses).  
XX  
SQ Sequence 18 AA;  
Query Match 31.2%; Score 5; DB 19; Length 18;  
Best Local Similarity 100.0%; Pred. No. 84;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 2 GYGRD 6  
Db 4 GYGRD 8  
RESULT 20  
AAG63536  
ID AAG63536 standard; Protein; 19 AA.  
XX  
XX AAG63536;  
XX  
XX 15-OCT-2001 (first entry)  
DT  
XX Amino acid sequence of a specific fragment of ATP1B isoform hATP1B.  
XX  
XX Human; ATP1; hATP1P2; hATP1P3; hATP1P4; hATP1P5; hATP1P6; AT2 receptor;  
KW angiotensin II receptor; antioncogenic; 8p21.3-p22; cancer.  
XX  
XX Homo sapiens.  
XX  
XX MO200157209-A2.  
XX  
XX 09-AUG-2001.  
PD  
XX 07-FEB-2001; 2001WO-FR00359.  
PF  
XX 07-FEB-2000; 2000FR-0001504.  
PR  
XX (CNRS ) CNRS CENT NAT RECH SCI.  
XX  
XX Nahmias C, Strosberg AD, Nouet S;  
XX  
XX WPI; 2001-488880/53.  
XX  
XX New protein family, designated hATP1, which interacts with the AT2  
PT receptor of angiotensin II are anti-oncogenic and useful to detect and  
PT treat cancer or precancerous conditions -  
XX  
XX Claim 16; Page 94; 118pp; French.  
XX  
XX The present sequence represents a fragment specific to an isoform of  
CC the human ATP protein, designated hATP1P2. ATP has isoforms designated  
CC hATP1P2, hATP1P3, hATP1P4, hATP1P5 and hATP1P6. All ATP proteins comprise  
CC their C-terminals a common fragment which interacts with the angiotensin  
CC II (AT2) receptor. ATP proteins have antioncogenic functions. The human  
CC ATP gene has 17 exons, and is located at chromosome region 8p21.3-p22.

CC ATRP polynucleotides and polypeptides are used to detect, evaluate or  
 CC give prognosis for a cancer or pre-cancer condition, and as an  
 CC anti-tumour medicament.

XX Sequence 19 AA;

Query Match 31.2%; Score 5; DB 22; Length 19;  
 Best Local Similarity 100.0%; Pred. No. 88;  
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 12 PSRKN 16  
 |||||

Db 8 PSRKN 12

RESULT 21

ID AAM50812 standard; Protein; 40 AA.

XX AAM50812;

DT 01-MAY-2002 (first entry)

XX PS118 prostate marker immunogenic polypeptide.

XX PS118; prostate; marker; prostate cancer; tumour; metastasis;  
 KW benign prostatic hyperplasia; prostatic intraepithelial neoplasia;  
 KW prostatitis; human; diagnosis; therapy; vaccine; immunogen.

OS Homo sapiens.

XX US2001055758-A1.

XX 27-DEC-2001.

XX 23-APR-1998; 98US-0065383.

XX 23-APR-1997; 97US-0842385.

PA (BILL/) BILLING-MEDEL P A.

PA (COHE/) COHEN M.

PA (COPL/) COPLPITTS T L.

PA (FRIE/) FRIEDMAN P N.

PA (GORD/) GORDON J.

PA (GRAN/) GRANADOS E N.

PA (HODG/) HODGES S C.

PA (KLAS/) KLAS M R.

PA (KRAT/) KRATOCHVIL J D.

PA (ROBE/) ROBERTS-RAPP L.

XX Billing-Medel PA, Cohen M, Coplitts TL, Friedman PN, Gordon J;  
 PI Granados EN, Hodges SC, Klasse MR, Kratochvil JD, Roberts-Rapp L;  
 PI Russell JC, Stroupe SD;

XX WPI; 2002-187683/24.

XX Claim 17; Page 44; 57pp; English.

CC The present sequence is that of an immunogenic polypeptide  
 CC comprising amino acids 332-371 of human prostate-specific PS118  
 CC polypeptide (see AAM50809). A PS118 consensus sequence (see  
 CC ABA51651) is found at least 12 times more often in prostate than  
 CC in non-prostate tissue. PS118 polypeptides, polynucleotides,  
 CC antibodies, agonists and inhibitors are useful for detecting,  
 CC diagnosing, staging, monitoring, prognosticating, preventing and  
 CC treating (including by genetic immunisation), or determining the  
 CC predisposition of an individual to, diseases and conditions of the  
 CC prostate, such as benign prostatic hyperplasia, prostatitis,

CC prostatic intraepithelial neoplasia, prostate cancer, tumours and  
 CC metastases. The PS118 polypeptides can be produced by expression  
 CC of PS118 polynucleotides in transfected host cells, and  
 CC immunogenic peptides are useful for raising PS118-specific  
 CC antibodies of diagnostic use. The methods and reagents of the  
 CC invention may provide an early means of detecting diseases of the  
 CC prostate and may also provide new markers which can differentiate  
 CC between the clinically important and unimportant prostate cancers  
 CC without the use of surgery.

XX Sequence 40 AA;

Query Match 31.2%; Score 5; DB 23; Length 40;  
 Best Local Similarity 100.0%; Pred. No. 1.7e+02;  
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 11 DPSRK 15  
 |||||

Db 8 DPSRK 12

RESULT 22

ID AAU39143 standard; Protein; 51 AA.

XX AAU39143;

DT 13-FEB-2002 (first entry)

XX Propionibacterium acnes immunogenic protein #39.

XX SAPHO syndrome; synovitis; acne; pustulosis; hypertosis; osteomyelitis;  
 KW uveitis; endophthalmitis; bone; joint; central nervous system; ELISA;  
 KW inflammatory lesion; acne vulgaris; enzyme linked immunosorbent assay;  
 KW dermatological; osteopathic; neuroprotectant.

OS Propionibacterium acnes.

XX WO200181581-A2.

XX 01-NOV-2001.

XX 20-APR-2001; 2001WO-US12865.

XX 21-APR-2000; 2000US-199047P.

XX 02-JUN-2000; 2000US-208841P.

XX 07-JUL-2000; 2000US-216747P.

XX (CORI-) CORIXA CORP.

XX Skeiky YAW, Persing DH, Mitcham JL, Wang SS, Bhattacha A;  
 PI L'maisonneuve J, Zhang Y, Jen S, Carter D;  
 PI WPI; 2001-616774/71.

XX DR N-PSDB; AAS59506.

XX Propionibacterium acnes polypeptides and nucleic acids useful for  
 PT vaccinating against and diagnosing infections, especially useful for  
 PT treating acne vulgaris -

XX Example 1; SEQ ID No 338; 1069pp; English.

XX Sequences AAU39105-AAU68017 represent Propionibacterium acnes immunogenic  
 CC polypeptides. The proteins and their associated DNA sequences are used in  
 CC the treatment, prevention and diagnosis of medical conditions caused by  
 CC P. acnes. The disorders include SAPHO syndrome (synovitis, acne,  
 CC pustulosis, hypertosis and osteomyelitis), uveitis and endophthalmitis.  
 CC P. acnes is also involved in infections of bone, joints and the central  
 CC nervous system, however it is particularly involved in the inflammatory  
 CC lesions associated with acne vulgaris. A method for detecting the  
 CC presence or absence of P. acnes in a patient comprises contacting a  
 CC sample with a binding agent that binds to the proteins of the invention  
 CC and determining the amount of bound protein in the sample. The

CC polypeptides may be used as antigens in the production of antibodies  
CC specific for *P. acnes* proteins. These antibodies can be used to  
CC downregulate expression and activity of *P. acnes* polypeptides and  
CC therefore treat *P. acnes* infections. The antibodies may also be used as  
CC diagnostic agents for determining *P. acnes* presence, for example, by  
CC enzyme linked immunosorbent assay (ELISA).  
CC Note: The sequence data for this patent did not form part of the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pct\_sequences.  
XX  
SQ Sequence 51 AA;  
Query Match 31.2%; Score 5; DB 22; Length 51;  
Best Local Similarity 100.0%; Pred. No. 2.1e+02;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 4 GRDDN 8  
DB 37 GRDDN 41  
RESULT 23  
AAU64366  
ID AAU64366 standard; Protein; 51 AA.  
XX  
AC AAU64366;  
XX  
DT 27-FEB-2002 (first entry)  
XX  
DE Propionibacterium acnes immunogenic protein #25262.  
XX  
KM SAPHO syndrome; synovitis; acne; pustulosis; hypertosis; osteomyelitis;  
KM uveitis; endophthalmitis; bone joint; central nervous system; ELISA;  
KM inflammatory lesion; acne vulgaris; enzyme linked immunosorbent assay;  
KM dermatological; osteopathic; neuroprotectant.  
XX  
OS Propionibacterium acnes.  
XX  
PN WO200181581-A2.  
XX  
PD 01-NOV-2001.  
XX  
PF 20-APR-2001; 2001WO-US12865.  
XX  
PR 21-APR-2000; 2000US-199047P.  
PR 02-JUN-2000; 2000US-208841P.  
PR 07-JUL-2000; 2000US-216747P.  
XX  
PA (CORI-) CORIXA CORP.  
XX  
PI Skeiky YAM, Persing DH, Mitcham JL, Wang SS, Bhatia A;  
PI L'maisonneuve J, Zhang Y, Jen S, Carter D;  
XX  
DR WPI; 2001-616774/71.  
DR N-PSDB; AAS59642.  
XX  
PT Propionibacterium acnes polypeptides and nucleic acids useful for  
PT vaccinating against and diagnosing infections, especially useful for  
PT treating acne vulgaris -  
XX  
PS Example 1; SEQ ID No 25561; 1069pp; English.  
XX  
CC Sequences AAU39105-AAU6017 represent Propionibacterium acnes immunogenic  
CC polypeptides. The proteins and their associated DNA sequences are used in  
CC the treatment, prevention and diagnosis of medical conditions caused by  
CC *P. acnes*. The disorders include SAPHO syndrome (synovitis, acne,  
CC pustulosis, hypertosis and osteomyelitis), uveitis and endophthalmitis.  
CC *P. acnes* is also involved in infections of bone, joints and the central  
CC nervous system, however it is particularly involved in the inflammatory  
CC lesions associated with acne vulgaris. A method for detecting the  
CC presence or absence of *P. acnes* in a patient comprises contacting a  
CC sample with a binding agent that binds to the proteins of the invention  
CC and determining the amount of bound protein in the sample. The

CC polypeptides may be used as antigens in the production of antibodies  
CC specific for *P. acnes* proteins. These antibodies can be used to  
CC downregulate expression and activity of *P. acnes* polypeptides and  
CC therefore treat *P. acnes* infections. The antibodies may also be used as  
CC diagnostic agents for determining *P. acnes* presence, for example, by  
CC enzyme linked immunosorbent assay (ELISA).  
CC Note: The sequence data for this patent did not form part of the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pct\_sequences.  
XX  
SQ Sequence 51 AA;  
Query Match 31.2%; Score 5; DB 22; Length 51;  
Best Local Similarity 100.0%; Pred. No. 2.1e+02;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 9 DRDPS 13  
DB 20 DRDPS 24  
RESULT 24  
AAU66009  
ID AAU66009 standard; Protein; 51 AA.  
XX  
AC AAU66009;  
XX  
DT 27-FEB-2002 (first entry)  
XX  
DE Propionibacterium acnes immunogenic protein #26905.  
XX  
KM SAPHO syndrome; synovitis; acne; pustulosis; hypertosis; osteomyelitis;  
KM uveitis; endophthalmitis; bone joint; central nervous system; ELISA;  
KM inflammatory lesion; acne vulgaris; enzyme linked immunosorbent assay;  
KM dermatological; osteopathic; neuroprotectant.  
XX  
OS Propionibacterium acnes.  
XX  
PN WO200181581-A2.  
XX  
PD 01-NOV-2001.  
XX  
PF 20-APR-2001; 2001WO-US12865.  
XX  
PR 21-APR-2000; 2000US-199047P.  
PR 02-JUN-2000; 2000US-208841P.  
PR 07-JUL-2000; 2000US-216747P.  
XX  
PA (CORI-) CORIXA CORP.  
XX  
PI Skeiky YAM, Persing DH, Mitcham JL, Wang SS, Bhatia A;  
PI L'maisonneuve J, Zhang Y, Jen S, Carter D;  
XX  
DR WPI; 2001-616774/71.  
DR N-PSDB; AAS59706.  
XX  
PT Propionibacterium acnes polypeptides and nucleic acids useful for  
PT vaccinating against and diagnosing infections, especially useful for  
PT treating acne vulgaris -  
XX  
PS Example 1; SEQ ID No 27204; 1069pp; English.  
XX  
CC Sequences AAU39105-AAU68017 represent Propionibacterium acnes immunogenic  
CC polypeptides. The proteins and their associated DNA sequences are used in  
CC the treatment, prevention and diagnosis of medical conditions caused by  
CC *P. acnes*. The disorders include SAPHO syndrome (synovitis, acne,  
CC pustulosis, hypertosis and osteomyelitis), uveitis and endophthalmitis.  
CC *P. acnes* is also involved in infections of bone, joints and the central  
CC nervous system, however it is particularly involved in the inflammatory  
CC lesions associated with acne vulgaris. A method for detecting the  
CC presence or absence of *P. acnes* in a patient comprises contacting a  
CC sample with a binding agent that binds to the proteins of the invention  
CC and determining the amount of bound protein in the sample. The

CC polypeptides may be used as antigens in the production of antibodies  
CC specific for P. acnes proteins. These antibodies can be used to  
CC downregulate expression and activity of P. acnes polypeptides and  
CC therefore treat P. acnes infections. The antibodies may also be used as  
CC diagnostic agents for determining P. acnes presence, for example, by  
CC enzyme linked immunosorbent assay (ELISA).  
CC Note: The sequence data for this patent did not form part of the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pct\_sequences.

SQ Sequence 51 AA;

Query Match 31.2%; Score 5; DB 22; Length 51;  
Best Local Similarity 100.0%; Pred. No. 2.1e+02;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 9 DRDPS 13  
DB 20 DRDPS 24

RESULT 25  
ABB27282  
ID ABB27282 standard; Protein; 51 AA.  
XX  
AC ABB27282;  
XX

DT 23-JAN-2002 (first entry)

XX Protein #9281 encoded by probe for measuring heart cell gene expression.

KW Human; gene expression; heart; microarray; vascular system;  
KM cardiovascular disease; hypertension; cardiac arrhythmia;  
XX congenital heart disease.

OS Homo sapiens.

PN WO200157274-A2.

PD 09-AUG-2001.

PF 30-JAN-2001; 2001WO-US00666.

XX 04-FEB-2000; 2000US-0180312.

PR 26-MAY-2000; 2000US-0207456.

PR 30-JUN-2000; 2000US-0608408.

PR 03-AUG-2000; 2000US-0632366.

PR 21-SEP-2000; 2000US-0234687.

PR 27-SEP-2000; 2000US-0236359.

PR 04-OCT-2000; 2000GB-0024263.

XX (MOLE-) MOLECULAR DYNAMICS INC.

PI Penn SG, Hanzel DK, Chen W, Rank DR;

DR WPI; 2001-488899/53.

XX Single exon nucleic acid probes for analyzing gene expression in human

XX hearts -

XX Claim 15; SEQ ID NO 29052; 530pp; English.

CC The present invention relates to single exon nucleic acid probes for  
CC measuring human gene expression in a sample derived from human heart (see  
CC ABA2155-ABA41305). The present sequence is a protein encoded by one such  
CC probe. The probes may be used for predicting, measuring and displaying  
CC gene expression in samples derived from the human heart via microarrays.  
CC By measuring gene expression, the probes are useful for predicting,  
CC diagnosing, grading, staging, monitoring and prognosing diseases of the  
CC human heart and vascular system e.g. cardiovascular disease,  
CC hypertension, cardiac arrhythmias and congenital heart disease.  
CC Note: The sequence data for this patent did not form part of the printed  
CC specification, but was obtained in electronic format directly from WIPO

CC at ftp.wipo.int/pub/published\_pct\_sequences.

XX SQ Sequence 51 AA;

Query Match 31.2%; Score 5; DB 22; Length 51;  
Best Local Similarity 100.0%; Pred. No. 2.1e+02;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 12 PSRKN 16  
DB 21 PSRKN 25

RESULT 26  
ABB40932  
ID ABB40932 standard; Peptide; 52 AA.  
XX  
AC ABB40932;  
XX

DT 04-FEB-2002 (first entry)

XX Peptide #8438 encoded by human foetal liver single exon probe.

KM Human; foetal liver; gene expression; single exon nucleic acid probe.

OS Homo sapiens.

PN WO200157277-A2.

PD 09-AUG-2001.

PF 30-JAN-2001; 2001WO-US00669.

XX 04-FEB-2000; 2000US-0180312.

PR 26-MAY-2000; 2000US-0207456.

PR 30-JUN-2000; 2000US-0608408.

PR 03-AUG-2000; 2000US-0632366.

PR 21-SEP-2000; 2000US-0234687.

PR 27-SEP-2000; 2000US-0236359.

PR 04-OCT-2000; 2000GB-0024263.

XX (MOLE-) MOLECULAR DYNAMICS INC.

PI Penn SG, Hanzel DK, Chen W, Rank DR;

DR WPI; 2001-483447/52.

XX Human genome-derived single exon nucleic acid probes useful for

XX analyzing gene expression in human fetal liver -

XX Claim 27; SEQ ID NO 33567; 639pp + sequence listing; English.

CC The invention relates to a single exon nucleic acid probe for  
CC measuring human gene expression in a sample derived from human foetal  
CC liver. The single exon nucleic acid probes may be used for predicting,  
CC measuring and displaying gene expression in samples derived from human  
CC foetal liver. The present sequence is a peptide encoded by a single exon  
CC nucleic acid probe of the invention.  
CC Note: The sequence data for this patent did not form part of the  
CC printed specification, but was obtained in electronic format directly  
CC from WIPO at ftp.wipo.int/pub/published\_pct\_sequences.

XX SQ Sequence 52 AA;

Query Match 31.2%; Score 5; DB 22; Length 52;  
Best Local Similarity 100.0%; Pred. No. 2.1e+02;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 12 PSRKN 16  
DB 38 PSRKN 42

```

RESULT 27
ID ABB25053 standard; Protein; 52 AA.
XX
AC ABB25053;
XX
DT 23-JAN-2002 (first entry)
XX
DE Protein #7052 encoded by probe for measuring heart cell gene expression.
XX
KW Human; gene expression; heart; microarray; vascular system;
KW cardiovascular disease; hypertension; cardiac arrhythmia;
KW congenital heart disease.
XX
OS Homo sapiens.
XX
PN WO200157274-A2.
XX
PD 09-AUG-2001.
XX
PF 30-JAN-2001; 2001WO-US00666.
XX
PR 04-FEB-2000; 2000US-0180312.
PR 26-MAY-2000; 2000US-0207456.
PR 30-JUN-2000; 2000US-0608408.
PR 03-AUG-2000; 2000US-0632366.
PR 21-SEP-2000; 2000US-0234687.
PR 27-SEP-2000; 2000US-0236359.
PR 04-OCT-2000; 2000GB-0024263.
XX
PA (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX
DR WPI; 2001-488899/53.
XX
PT Single exon nucleic acid probes for analyzing gene expression in human
PT hearts -
XX
PS Claim 15; SEQ ID NO 26823; 530bp; English.
XX
CC The present invention relates to single exon nucleic acid probes for
CC measuring human gene expression in a sample derived from human heart (see
CC ABA21535-BA41305). The present sequence is a protein encoded by one such
CC probe. The probes may be used for predicting, measuring and displaying
CC gene expression in samples derived from the human heart via microarrays.
CC By measuring gene expression, the probes are useful for predicting,
CC diagnosing, grading, staging, monitoring and prognosing diseases of the
CC human heart and vascular system e.g. cardiovascular disease,
CC hypertension, cardiac arrhythmias and congenital heart disease.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 52 AA;
XX
Query Match 31.2%; Score 5; DB 22; Length 52;
Best Local Similarity 100.0%; Pred. No. 2.1e+02;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 12 PSRKN 16
DB 38 PSRKN 42
XX
RESULT 28
ID AAM61794 standard; Protein; 52 AA.
XX
AC AAM61794;
XX
DT 05-NOV-2001 (first entry)
XX

```

```

DE Human brain expressed single exon probe encoded protein SEQ ID NO: 33899.
XX
KW Human; brain expressed exon; gene expression analysis; probe;
KW microarray; Alzheimer's disease; multiple sclerosis; schizophrenia;
KW epilepsy; cancer.
XX
OS Homo sapiens.
XX
PN WO200157275-A2.
XX
PD 09-AUG-2001.
XX
PF 30-JAN-2001; 2001WO-US00667.
XX
PR 04-FEB-2000; 2000US-0180312.
PR 26-MAY-2000; 2000US-0207456.
PR 30-JUN-2000; 2000US-0608408.
PR 03-AUG-2000; 2000US-0632366.
PR 21-SEP-2000; 2000US-0234687.
PR 27-SEP-2000; 2000US-0236359.
PR 04-OCT-2000; 2000GB-0024263.
XX
PA (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX
DR WPI; 2001-483446/52.
XX
PT Single exon nucleic acid probes for analyzing gene expression in human
PT brains -
XX
PS Example 4; SEQ ID NO: 33899; 650bp + Sequence Listing; English.
XX
CC The present invention provides a number of single exon nucleic acid
CC probes which are derived from genomic sequences expressed in the human
CC brain. They can be used to measure gene expression in brain cell samples,
CC which may enable the diagnosis and improved treatment of nervous system
CC diseases such as Alzheimer's disease, multiple sclerosis, schizophrenia,
CC epilepsy and cancers. The present sequence is a protein encoded by one of
CC the probes of the invention.
XX
SQ Sequence 52 AA;
XX
Query Match 31.2%; Score 5; DB 22; Length 52;
Best Local Similarity 100.0%; Pred. No. 2.1e+02;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 12 PSRKN 16
DB 38 PSRKN 42
XX
RESULT 29
ID AAM74593 standard; Protein; 52 AA.
XX
AC AAM74593;
XX
DT 06-NOV-2001 (first entry)
XX
DE Human bone marrow expressed probe encoded protein SEQ ID NO: 34899.
XX
KW Human; bone marrow expressed exon; gene expression analysis; probe;
KW microarray; cancer; leukaemia; lymphoma; myeloma.
XX
OS Homo sapiens.
XX
PN WO200157276-A2.
XX
PD 09-AUG-2001.
XX
PF 30-JAN-2001; 2001WO-US00668.
XX

```

PR 04-FEB-2000; 2000US-0180312.  
PR 26-MAY-2000; 2000US-0207456.  
PR 30-JUN-2000; 2000US-0608408.  
PR 03-AUG-2000; 2000US-0632366.  
PR 21-SEP-2000; 2000US-0234687.  
PR 27-SEP-2000; 2000US-0236359.  
PR 04-OCT-2000; 2000GB-0024263.  
XX  
XX (MOLE-) MOLECULAR DYNAMICS INC.  
XX  
XX Penn SG, Hanzel DK, Chen W, Rank DR;  
XX  
XX WPI; 2001-488900/53.  
XX  
XX Human genome-derived single exon nucleic acid probes useful for  
XX analyzing gene expression in human bone marrow -  
XX  
XX Example 4; SEQ ID No: 34899; 658pp + Sequence Listing; English.  
XX  
XX The present invention provides a number of single exon nucleic acid  
XX probes which are derived from genomic sequences expressed in the human  
XX bone marrow. They can be used to measure gene expression in bone marrow  
XX samples, which may enable the improved diagnosis and treatment of cancers  
XX such as lymphoma, leukemia and myeloma. The present sequence is a  
XX protein encoded by one of the probes of the invention.  
XX  
XX Sequence 52 AA;  
SQ

Query Match 31.2%; Score 5; DB 22; Length 52;  
Best Local Similarity 100.0%; Pred. No. 2.1e+02;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 12 PSRKN 16  
|||  
38 PSRKN 42

DB

RESULT 30  
AAM20353  
ID AAM20353 standard; Protein; 52 AA.  
XX  
XX AAM20353;  
XX

DT 12-OCT-2001 (first entry)

DE Peptide #6787 encoded by probe for measuring cervical gene expression.  
XX  
XX Probe; human; microarray; gene expression; cervical epithelial cell;  
XX cervical cancer.  
XX

OS Homo sapiens.

XX WO200157278-A2.

XX 09-AUG-2001.

PD 30-JAN-2001; 2001WO-US00670.

XX 04-FEB-2000; 2000US-0180312.

PR 26-MAY-2000; 2000US-0207456.

PR 30-JUN-2000; 2000US-0608408.

PR 03-AUG-2000; 2000US-0632366.

PR 21-SEP-2000; 2000US-0234687.

PR 27-SEP-2000; 2000US-0236359.

PR 04-OCT-2000; 2000GB-0024263.

XX (MOLE-) MOLECULAR DYNAMICS INC.  
XX  
XX Penn SG, Hanzel DK, Chen W, Rank DR;  
XX  
XX WPI; 2001-488901/53.  
XX  
XX Human genome-derived single exon nucleic acid probes useful for

PT analyzing gene expression in human cervical epithelial cells -  
XX  
XX Claim 27; SEQ ID No 25179; 487pp; English.  
XX

CC The present invention relates to human single exon nucleic acid probes  
CC (SENPs: see AAI10068-AA128459). The present sequence is a peptide encoded  
CC by one such probe. The SENPs are derived from human Hela cells. The SENPs  
CC can be used to produce a single exon microarray, which can be used for  
CC measuring human gene expression in a sample derived from human cervical  
CC epithelial cells. By measuring gene expression, the probes are therefore  
CC useful in grading and/or staging of diseases of the cervix, notably  
CC cervical cancer.  
CC Note: The sequence data for this patent did not form part of the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pct\_sequences.  
XX

SQ Sequence 52 AA;

Query Match 31.2%; Score 5; DB 22; Length 52;  
Best Local Similarity 100.0%; Pred. No. 2.1e+02;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 12 PSRKN 16  
|||  
38 PSRKN 42

DB

RESULT 31  
AAM34705  
ID AAM34705 standard; Protein; 52 AA.  
XX  
XX AAM34705;  
XX

DT 17-OCT-2001 (first entry)

DE Peptide #8742 encoded by probe for measuring placental gene expression.

XX Probe; microarray; human; placenta; antenatal diagnosis;  
XX genetic disorder.  
XX

OS Homo sapiens.

XX WO200157272-A2.

XX 09-AUG-2001.

PD 30-JAN-2001; 2001WO-US00663.

XX 04-FEB-2000; 2000US-0180312.

PR 26-MAY-2000; 2000US-0207456.

PR 30-JUN-2000; 2000US-0608408.

PR 03-AUG-2000; 2000US-0632366.

PR 21-SEP-2000; 2000US-0234687.

PR 27-SEP-2000; 2000US-0236359.

PR 04-OCT-2000; 2000GB-0024263.

XX (MOLE-) MOLECULAR DYNAMICS INC.  
XX  
XX Penn SG, Hanzel DK, Chen W, Rank DR;  
XX  
XX WPI; 2001-48897/53.  
XX

PT Human genome-derived single exon nucleic acid probes useful for  
XX analyzing gene expression in human placenta -  
XX  
XX Claim 27; SEQ ID No 34974; 654pp; English.  
XX

CC The present invention relates to single exon nucleic acid probes (SENPs:  
CC see AAI31315-AA157546). The present sequence is a peptide encoded by one  
CC such probe. The probes are useful for producing a microarray for  
CC predicting, measuring and displaying gene expression in samples derived  
CC from human placenta. The probes are useful for antenatal diagnosis of  
CC human genetic disorders.

XX Sequence 52 AA;  
 Query Match 31.2%; Score 5; DB 22; Length 52;  
 Best Local Similarity 100.0%; Pred. No. 2.1e+02;  
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 12 PSRXN 16  
 Db 38 PSRXN 42  
 RESULT 32  
 ABG44431  
 ID ABG44431 standard; Peptide; 52 AA.  
 XX  
 AC ABG44431;  
 XX  
 DT 19-AUG-2002 (first entry)  
 XX  
 DE Human peptide encoded by genome-derived single exon probe SEQ ID 34096.  
 XX  
 KW Human; single exon probe; asthma; lung cancer; COPD; ILD;  
 KW chronic obstructive pulmonary disease; interstitial lung disease;  
 KW familial idiopathic pulmonary fibrosis; neurofibromatosis;  
 KW tuberous sclerosis; Gaucher's disease; Niemann-Pick disease;  
 KW Hermansky-Pudlak syndrome; sarcoidosis; pulmonary haemosiderosis;  
 KW pulmonary histiocytosis; lymphangioleiomyomatosis; Kargener syndrome;  
 KW pulmonary alveolar proteinosis; fibrocystic pulmonary dysplasia;  
 KW primary ciliary dyskinesia; pulmonary hypertension;  
 KW hyaline membrane disease.  
 XX  
 XX Homo sapiens.  
 OS  
 PN WO200186003-A2.  
 PD 15-NOV-2001.  
 XX  
 PF 30-JAN-2001; 2001WO-US00665.  
 XX  
 PR 04-FEB-2000; 2000US-180312P.  
 PR 26-MAY-2000; 2000US-207456P.  
 PR 30-JUN-2000; 2000US-0608408.  
 PR 03-AUG-2000; 2000US-0632366.  
 PR 21-SEP-2000; 2000US-234687P.  
 PR 27-SEP-2000; 2000US-236359P.  
 PR 04-OCT-2000; 2000GB-0024263.  
 XX  
 PA (MOLE-) MOLECULAR DYNAMICS INC.  
 PI Penn SG, Hanzel DK, Chen W, Rank DR;  
 XX  
 DR WPI; 2002-114183/15.  
 XX  
 PT Spatially-addressable set of single exon nucleic acid probes, used to  
 measure gene expression in human lung samples -  
 XX  
 PS Claim 27; SEQ ID No 34096; 634pp; English.  
 XX  
 CC The invention relates to a spatially-addressable set of single exon  
 CC nucleic acid probes for measuring gene expression in a sample derived  
 CC from human lung comprising single exon nucleic acid probes having one of  
 CC 12614 nucleic acid sequences mentioned in the specification, or their  
 CC complements or the 12387 open reading frames derived from the 12614  
 CC probes. Also included are a microarray comprising the novel set of  
 CC probes; the novel set of probes which hybridize at high stringency to a  
 CC nucleic acid expressed in the human lung; measuring gene expression in a  
 CC sample derived from human lung, comprising (a) contacting the array with a  
 CC collection of detectably labeled nucleic acids derived from human lung  
 CC mRNA, and (b) measuring the label detectably bound to each probe of  
 CC the array; identifying exons in a eukaryotic genome, comprising  
 CC (a) algorithmically predicting at least one exon from genomic sequences  
 CC of the eukaryote; and (b) detecting specific hybridisation of detectably

CC labeled nucleic acids from eukaryote lung mRNA, to a single exon probe,  
 CC having a fragment identical to the predicted exon, the probe is included  
 CC in the above mentioned microarray; assigning exons to a single gene,  
 CC comprising (a) identifying exons from genomic sequence by the method  
 CC above and (b) measuring the expression of each of the exons in several  
 CC tissues and/or cell types using hybridisation to a single exon  
 CC microarray having a probe with the exon, where a common pattern of  
 CC expression of the exons in the tissues and/or cell types indicates that  
 CC the exons should be assigned to a single gene; a peptide comprising one  
 CC of 12011 sequences, mentioned in the specification, or encoded by the  
 CC probes/open reading frames (ORF). The probes are used for gene  
 CC expression analysis, and for identifying exons in a gene, particularly  
 CC using human lung derived mRNA and for the study of lung diseases  
 CC such as asthma, lung cancer, chronic obstructive pulmonary disease  
 CC (COPD), interstitial lung disease (ILD), familial idiopathic pulmonary  
 CC fibrosis, neurofibromatosis, tuberous sclerosis, Gaucher's disease,  
 CC Niemann-Pick disease, Hermansky-Pudlak syndrome, sarcoidosis, pulmonary  
 CC haemosiderosis, pulmonary histiocytosis, lymphangioleiomyomatosis,  
 CC pulmonary alveolar proteinosis, Kargener syndrome, fibrocystic  
 CC pulmonary dysplasia, primary ciliary dyskinesia, pulmonary hypertension  
 CC and hyaline membrane disease. The present sequence is a peptide/protein  
 CC encoded by a single exon probe of the invention.  
 CC Note: The sequence data for this patent did not form part  
 CC of the printed specification, but was obtained in electronic  
 CC format directly from WIPO at  
 CC ftp.wipo.int/pub/published\_pct\_sequences.  
 XX  
 SQ Sequence 52 AA;  
 Query Match 31.2%; Score 5; DB 23; Length 52;  
 Best Local Similarity 100.0%; Pred. No. 2.1e+02;  
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 12 PSRXN 16  
 Db 38 PSRXN 42  
 RESULT 33  
 ABP32718  
 ID ABP32718 standard; Protein; 54 AA.  
 XX  
 AC ABP32718;  
 XX  
 DT 08-JUL-2002 (first entry)  
 XX  
 DE Human ORF1691 protein, SEQ ID NO:3382.  
 XX  
 KW Human; ORF; open reading frame; ORF; drug screening; diagnosis;  
 KW disease monitoring; cytokine; cell proliferation; cell differentiation;  
 KW immune modulation; haematopoiesis regulation; tissue growth;  
 KW angiogenesis; activin; inhibin; chemotactic; chemokinesis; haemostatic;  
 KW thrombolytic; tumour inhibition; bodily characteristics; fertility;  
 KW behaviour; cancer; proliferative disorder; neurological disorder;  
 KW cardiovascular disease; immune system disorder; organ transplantation;  
 KW tissue growth disorder; tissue regeneration disorder; diabetes mellitus;  
 KW hypothyroidism; cholesterol ester storage disease; infection; vulvar;  
 KW vasotropic; antipsoriatic; antidiabetic; cytostatic; nocotropic;  
 KW neuroprotective; antihypertensive; antitumor; antidiabetic; thrombolytic;  
 KW cardiant; hypotensive; antithyroid; antiinflammatory; immunomodulator;  
 KW dermatological; analgesic; virucide; antibacterial; fungicide.  
 XX  
 OS Homo sapiens.  
 PN WO200190366-A2.  
 PD 29-NOV-2001.  
 XX  
 PE 24-MAY-2001; 2001WO-US17076.  
 XX  
 PR 24-MAY-2000; 2000US-206690P.  
 XX  
 PA (CURA-) CURAGEN CORP.

XX Leach MD, Shinkets RA;  
 PI  
 XX  
 DR MPI: 2002-106200/14.  
 DR N-PSDB; ABN76744.  
 XX  
 PT Novel human polypeptides and polynucleotides useful for diagnosing,  
 PT preventing and treating cardiovascular disease, neurodegenerative,  
 PT hyperproliferative disorders and disorders related to organ  
 PT transplantation -  
 PS  
 XX Claim 10; Page 1105; 2508bp; English.  
 PS  
 CC Sequences ABP31028-ABP35561 represent 4534 novel human proteins  
 CC designated ORF (open reading frame) 1-4534, and sequences ABN75054-  
 CC ABN79587 represent cDNAs encoding them. The invention also encompasses  
 CC polypeptides at least 80% identical to the ORF1-ORF4534 (collectively  
 CC referred to as ORFX) proteins, polynucleotides at least 85% identical to  
 CC the ORFX nucleic acid sequences, vectors and host cells comprising ORFX  
 CC polynucleotides, the recombinant production of ORFX proteins, antibodies  
 CC specific for ORFX proteins, methods of detecting ORFX polynucleotides and  
 CC polypeptides, methods of screening for modulators of ORFX expression or  
 CC activity, and methods of screening individuals for a predisposition to an  
 CC ORFX-associated disorder. The ORFX proteins of the invention have a wide  
 CC range of biological activities, such as cytokine, cell proliferation,  
 CC cell differentiation, immune modulation, haematopoiesis regulation,  
 CC tissue growth, angiogenesis, activin or inhibin activity, chemotactic/  
 CC chemokinetic activity, haemostatic activity, thrombolytic activity,  
 CC receptor/ligand, antiinflammatory activity, tumour inhibition activity,  
 CC and antiinfective activity, and may also be involved in the determination  
 CC of bodily characteristics, fertility and behaviour. ORFX proteins,  
 CC nucleic acids and antibodies may be used in the treatment of cancers,  
 CC other proliferative disorders such as psoriasis and benign tumours,  
 CC neurological disorders such as epilepsy and Alzheimer's disease,  
 CC cardiovascular diseases, immune system disorders, disorders related to  
 CC organ transplantation, disorders of tissue growth and regeneration,  
 CC diseases such as diabetes mellitus, hypothyroidism, and cholesterol ester  
 CC storage disease, and infectious diseases caused by viral, bacterial,  
 CC fungal and other pathogens. ORFX nucleic acids may also be used as a  
 CC source of primers and probes, in the detection of ORFX genomic sequences  
 CC or transcripts, in the identification and cloning of homologous  
 CC sequences, in genetic diagnosis, and in forensic biology. The ORFX  
 CC nucleic acids may additionally be used to produce transgenic animals  
 CC which may be useful for studying the function and/or activity of ORFX  
 CC protein, and in drug screening. The ORFX proteins may also be used as  
 CC immunogens to generate specific antibodies, which are useful in the  
 CC diagnosis, treatment and monitoring of ORFX-associated diseases.  
 CC  
 XX  
 SQ Sequence 54 AA;  
 Query Match 31.2%; Score 5; DB 23; Length 54;  
 Best Local Similarity 100.0%; Pred. No. 2.2e+02;  
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 OY 10 RDPSP 14  
 |||||  
 Db 50 RDPSP 54  
 RESULT 34  
 AAU40965  
 ID AAU40965 standard; Protein; 57 AA.  
 XX  
 AC AAU40965;  
 XX  
 DT 13-FEB-2002 (first entry)  
 XX  
 DE Propionibacterium acnes immunogenic protein #1861.  
 XX  
 KW SAPHO syndrome; synovitis; acne; pustulosis; hyperostosis; osteomyelitis;  
 KW uveitis; endophthalmitis; bone; joint; central nervous system; ELISA;  
 KW inflammatory lesion; acne vulgaris; enzyme linked immunosorbent assay;  
 KW dermatological; osteopathic; neuroprotectant.

XX  
 OS Propionibacterium acnes.  
 XX  
 PN WO200181581-A2.  
 XX  
 XX  
 PD 01-NOV-2001.  
 XX  
 PF 20-APR-2001; 2001WO-US12865.  
 XX  
 PR 21-APR-2000; 2000US-199047P.  
 PR 02-JUN-2000; 2000US-208841P.  
 PR 07-JUL-2000; 2000US-216747P.  
 XX  
 PA (CORI-) CORIXA CORP.  
 XX  
 PI Skeiky YAM, Persing DH, Mitcham Jr., Wang SS, Bhalaria A;  
 PI L'maisonneuve J, Zhang Y, Jen S, Carter D;  
 DR MPI: 2001-616774/71.  
 DR N-PSDB; AAS59513.  
 XX  
 PS Example 1; SEQ ID No 2160; 1069bp; English.  
 XX  
 CC Sequences AAU39105-AAU68017 represent Propionibacterium acnes immunogenic  
 CC polypeptides. The proteins and their associated DNA sequences are used in  
 CC the treatment, prevention and diagnosis of medical conditions caused by  
 CC P. acnes. The disorders include SAPHO syndrome (synovitis, acne,  
 CC pustulosis, hyperostosis and osteomyelitis), uveitis and endophthalmitis.  
 CC P. acnes is also involved in infections of bone, joints and the central  
 CC nervous system, however it is particularly involved in the inflammatory  
 CC lesions associated with acne vulgaris. A method for detecting the  
 CC presence or absence of P. acnes in a patient comprises contacting a  
 CC sample with a binding agent that binds to the proteins of the invention  
 CC and determining the amount of bound protein in the sample. The  
 CC polypeptides may be used as antigens in the production of antibodies  
 CC specific for P. acnes proteins. These antibodies can be used to  
 CC downregulate expression and activity of P. acnes polypeptides and  
 CC therefore treat P. acnes infections. The antibodies may also be used as  
 CC diagnostic agents for determining P. acnes presence, for example, by  
 CC enzyme linked immunosorbent assay (ELISA).  
 CC Note: The sequence data for this patent did not form part of the printed  
 CC specification, but was obtained in electronic format directly from WIPO  
 CC at ftp.wipo.int/pub/published\_pct\_sequences.  
 CC  
 XX  
 SQ Sequence 57 AA;  
 Query Match 31.2%; Score 5; DB 22; Length 57;  
 Best Local Similarity 100.0%; Pred. No. 2.3e+02;  
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 OY 2 GYGRD 6  
 |||||  
 Db 9 GYGRD 13  
 RESULT 35  
 ABG26481  
 ID ABG26481 standard; Protein; 62 AA.  
 XX  
 AC ABG26481;  
 XX  
 DT 18-FEB-2002 (first entry)  
 XX  
 DE Novel human diagnostic protein #26472.  
 XX  
 KW Human; chromosome mapping; gene mapping; gene therapy; forensic;  
 KW food supplement; medical imaging; diagnostic; genetic disorder.  
 XX  
 OS Homo sapiens.



XX WO200175067-A2.  
PN 11-OCT-2001.  
XX  
XX 30-MAR-2001; 2001WO-US08631.  
PF  
XX 31-MAR-2000; 2000US-0540217.  
PR 23-AUG-2000; 2000US-0649167.  
XX  
PA (HYSE-) HYSEQ INC.  
XX  
PI Drmanac RT, Liu C, Tang YT;  
XX WPI; 2001-639362/73.  
DR N-PSDB; AAS90668.  
XX  
XX New isolated polynucleotide and encoded polypeptides, useful in  
PT diagnostics, forensics, gene mapping, identification of mutations  
PT responsible for genetic disorders or other traits and to assess  
PT biodiversity -  
XX  
PS Claim 20; SEQ ID No 56840; 103pp; English.  
XX  
XX The invention relates to isolated polynucleotide (I) and  
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,  
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome  
CC and gene mapping, and in recombinant production of (II). The  
CC polynucleotides are also used in diagnostics as expressed sequence tags  
CC for identifying expressed genes. (I) is useful in gene therapy techniques  
CC to restore normal activity of (II) or to treat disease states involving  
CC (II). (II) is useful for generating antibodies against it, detecting or  
CC quantitating a polypeptide in tissue, as molecular weight markers and as  
CC a food supplement. (II) and its binding partners are useful in medical  
CC imaging of sites expressing (II). (I) and (II) are useful for treating  
CC disorders involving aberrant protein expression or biological activity.  
CC The polypeptide and polynucleotide sequences have applications in  
CC diagnostics, forensics, gene mapping, identification of mutations  
CC responsible for genetic disorders or other traits to assess biodiversity  
CC and to produce other types of data and products dependent on DNA and  
CC amino acid sequences. ABG00010-ABG30377 represent novel human  
CC diagnostic amino acid sequences of the invention.  
CC Note: The sequence data for this patent did not appear in the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pct\_sequences.  
XX  
SQ Sequence 62 AA;  
Query Match 31.2%; Score 5; DB 22; Length 62;  
Best Local Similarity 100.0%; Pred. No. 2.5e+02;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 4 GRDDN 8  
|||  
Db 10 GRDDN 14  
|||  
RESULT 36  
ABP05665  
ID ABP05665 standard; Protein; 63 AA.  
XX  
AC ABP05665;  
XX  
DT 24-JUN-2002 (first entry)  
XX  
DE Human ORFX protein sequence SEQ ID NO:11312.  
XX  
XX Human; open reading frame; ORFX; gene therapy; cancer; cirrhosis;  
KM hyperproliferative disorder; psoriasis; benign tumour; hemorrhage;  
KM degenerative disorder; osteoarthritis; neurodegenerative disorder;  
KM cardiovascular disease; diabetes mellitus; systemic lupus erythematosus;  
KM hypertension; hypothyroidism; cholesterol ester storage disease;  
KM immune deficiency; immune disorder; infectious disease;

KM autoimmune disorder; rheumatoid arthritis; autoimmune thyroiditis;  
KM myasthenia gravis.  
XX  
XX Homo sapiens.  
OS  
XX WO200192523-A2.  
PN  
XX 06-DEC-2001.  
XX  
PD 29-MAY-2001; 2001WO-US10836.  
XX  
PF 30-MAY-2000; 2000US-206132P.  
PR 29-AUG-2000; 2000US-228716P.  
XX  
XX (CURA-) CURAGEN CORP.  
PA  
PI Shinkets RA, Leach MD;  
XX WPI; 2002-106308/14.  
DR N-PSDB; ABN21417.  
XX  
XX Novel human polypeptides and polynucleotides useful for diagnosing,  
PT preventing and treating cardiovascular disease, neurodegenerative,  
PT hyperproliferative disorders and autoimmune disorders -  
XX  
PS Disclosure; SEQ ID 11312; 1037pp; English.  
XX  
XX The present invention describes substantially purified human proteins  
CC (referred to as open reading frame, ORFX, where X is 1-11491 (see Table 1  
CC in the specification). ABN15762 to ABN27252 encode the human ORFX  
CC proteins given in ABP00010 to ABP11500. ORFX proteins are useful for  
CC treating or preventing a pathology associated with an ORFX-associated  
CC disorder in humans, and in the manufacture of a medicament for treating a  
CC syndrome associated with ORFX-associated disorder. ORFX polynucleotide  
CC sequences can be used in gene therapy. ORFX sequences can be used in the  
CC treatment of cancer, hyperproliferative disorders, cirrhosis of liver,  
CC psoriasis, benign tumours, keloid, degenerative disorders, haemorrhage,  
CC osteoarthritis, neurodegenerative disorders, disorders related to organ  
CC transplantation, cardiovascular diseases, diabetes mellitus, systemic  
CC lupus erythematosus, hypertension, hypothyroidism, cholesterol ester  
CC storage disease, various immune deficiencies and disorders, infectious  
CC diseases, autoimmune disorders such as multiple sclerosis, rheumatoid  
CC arthritis, autoimmune thyroiditis, myasthenia gravis, graft-versus-host  
CC disease and autoimmune inflammatory eye disease. ORFX proteins are also  
CC useful for treating burns, incisions, ulcers, for treating osteoporosis,  
CC bone degenerative disorders, or periodontal disease, and for gut,  
CC protection or regeneration and treatment of lung or liver fibrosis,  
CC reperfusion injury in various tissues and conditions resulting from  
CC systemic cytokine damage.  
CC N.B. The sequence data for this patent did not form part of the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pct\_sequences.  
XX  
SQ Sequence 63 AA;  
Query Match 31.2%; Score 5; DB 23; Length 63;  
Best Local Similarity 100.0%; Pred. No. 2.5e+02;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 10 RDPSP 14  
|||  
Db 1 RDPSP 5  
|||  
RESULT 37  
ABB39553  
ID ABB39553 standard; Peptide; 68 AA.  
XX  
AC ABB39553;  
XX  
DT 04-FEB-2002 (first entry)  
XX  
DE Peptide #7059 encoded by human foetal liver single exon probe.

```
XX Human; foetal liver; gene expression; single exon nucleic acid probe.
KM Homo sapiens.
OS WO200157277-A2.
XX
XX 09-AUG-2001.
XX
XX 30-JAN-2001; 2001WO-US00669.
XX
XX 04-FEB-2000; 2000US-0180312.
XX 26-MAY-2000; 2000US-0207456.
XX 30-JUN-2000; 2000US-0608408.
XX 03-AUG-2000; 2000US-0632366.
XX 21-SEP-2000; 2000US-0234687.
XX 27-SEP-2000; 2000US-0236359.
XX 04-OCT-2000; 2000GB-0024263.
XX
XX (MOLE-) MOLECULAR DYNAMICS INC.
XX
XX Penn SG, Hanzel DK, Chen W, Rank DR;
XX WPI; 2001-483447/52.
XX
XX Human genome-derived single exon nucleic acid probes useful for
XX analyzing gene expression in human fetal liver -
XX
XX Claim 27; SEQ ID NO 32188; 639pp + sequence listing; English.
XX
XX The invention relates to a single exon nucleic acid probe for
XX measuring human gene expression in a sample derived from human foetal
XX liver. The single exon nucleic acid probes may be used for predicting,
XX measuring and displaying gene expression in samples derived from human
XX foetal liver. The present sequence is a peptide encoded by a single exon
XX nucleic acid probe of the invention.
XX Note: The sequence data for this patent did not form part of the
XX printed specification, but was obtained in electronic format directly
XX from WIPO at ftp.wipo.int/pub/published_pct_sequences.
XX
XX Sequence 68 AA;
SQ
Query Match 31.2%; Score 5; DB 22; Length 68;
Best Local Similarity 100.0%; Pred. No. 2.7e+02;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 3 YGRDD 7
Db 40 YGRDD 44
RESULT 38
AAM60252
ID AAM60252 standard; Protein; 68 AA.
XX
XX AAM60252;
XX
XX 05-NOV-2001 (first entry)
XX
XX Human brain expressed single exon probe encoded protein SEQ ID NO: 32357.
DE
XX Human; brain expressed exon; gene expression analysis; probe;
KM microarray; Alzheimer's disease; multiple sclerosis; schizophrenia;
KW epilepsy; cancer.
XX
XX Homo sapiens.
OS
XX WO200157275-A2.
XX
XX 09-AUG-2001.
XX
XX 30-JAN-2001; 2001WO-US00667.
XX
```

```
PR 04-FEB-2000; 2000US-0180312.
PR 26-MAY-2000; 2000US-0207456.
PR 30-JUN-2000; 2000US-0608408.
PR 03-AUG-2000; 2000US-0632366.
PR 21-SEP-2000; 2000US-0234687.
PR 27-SEP-2000; 2000US-0236359.
PR 04-OCT-2000; 2000GB-0024263.
XX
XX (MOLE-) MOLECULAR DYNAMICS INC.
XX
XX Penn SG, Hanzel DK, Chen W, Rank DR;
XX WPI; 2001-483446/52.
XX
XX Single exon nucleic acid probes for analyzing gene expression in human
XX brains -
XX
XX Example 4; SEQ ID NO: 32357; 650pp + Sequence Listing; English.
XX
XX The present invention provides a number of single exon nucleic acid
XX probes which are derived from genomic sequences expressed in the human
XX brain. They can be used to measure gene expression in brain cell samples,
XX which may enable the diagnosis and improved treatment of nervous system
XX diseases such as Alzheimer's disease, multiple sclerosis, schizophrenia,
XX epilepsy and cancers. The present sequence is a protein encoded by one of
XX the probes of the invention.
XX
XX Sequence 68 AA;
SQ
Query Match 31.2%; Score 5; DB 22; Length 68;
Best Local Similarity 100.0%; Pred. No. 2.7e+02;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 3 YGRDD 7
Db 40 YGRDD 44
RESULT 39
AAM72879
ID AAM72879 standard; Protein; 68 AA.
XX
XX AAM72879;
XX
XX 06-NOV-2001 (first entry)
XX
XX Human bone marrow expressed probe encoded protein SEQ ID NO: 33185.
DE
XX Human; bone marrow expressed exon; gene expression analysis; probe;
KM microarray; cancer; leukaemia; lymphoma; myeloma.
XX
XX Homo sapiens.
OS
XX WO200157276-A2.
XX
XX 09-AUG-2001.
XX
XX 30-JAN-2001; 2001WO-US00668.
XX
XX 04-FEB-2000; 2000US-0180312.
XX 26-MAY-2000; 2000US-0207456.
XX 30-JUN-2000; 2000US-0608408.
XX 03-AUG-2000; 2000US-0632366.
XX 21-SEP-2000; 2000US-0234687.
XX 27-SEP-2000; 2000US-0236359.
XX 04-OCT-2000; 2000GB-0024263.
XX
XX (MOLE-) MOLECULAR DYNAMICS INC.
XX
XX Penn SG, Hanzel DK, Chen W, Rank DR;
XX WPI; 2001-488900/53.
XX
```

PT Human genome-derived single exon nucleic acid probes useful for  
XX analyzing gene expression in human bone marrow -  
PS Example 4; SEQ ID NO: 33185; 658pp + Sequence Listing; English.  
XX  
CC The present invention provides a number of single exon nucleic acid  
CC probes which are derived from genomic sequences expressed in the human  
CC bone marrow. They can be used to measure gene expression in bone marrow  
CC samples, which may enable the improved diagnosis and treatment of cancers  
CC such as lymphoma, leukemia and myeloma. The present sequence is a  
CC protein encoded by one of the probes of the invention.

SQ Sequence 68 AA;

Query Match

31.2%; Score 5; DB 22; Length 68;

Best Local Similarity 100.0%; Pred. No. 2.7e+02;

Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 YGRDD 7  
|||

Db 40 YGRDD 44

RESULT 40

AAM33110  
ID AAM33110 standard; Protein; 68 AA.

XX  
AC AAM33110;

DT 17-OCT-2001 (first entry)

XX Peptide #7147 encoded by probe for measuring placental gene expression.

DE Probe: microarray; human; placenta; antenatal diagnosis;

KM genetic disorder.

XX Homo sapiens.

OS  
PN WO200157272-A2.

XX  
PD 09-AUG-2001.

XX 30-JAN-2001; 2001WO-US00663.

XX 04-FEB-2000; 2000US-0180312.

XX 26-MAY-2000; 2000US-0207456.

XX 30-JUN-2000; 2000US-0608408.

XX 03-AUG-2000; 2000US-0632366.

XX 21-SEP-2000; 2000US-0234687.

XX 27-SEP-2000; 2000US-0236359.

XX 04-OCT-2000; 2000GB-0024263.

XX (MOLE-) MOLECULAR DYNAMICS INC.

XX Penn SG, Hanzel DK, Chen W, Rank DR;

XX WPI; 2001-488897/53.

XX Human genome-derived single exon nucleic acid probes useful for

XX analyzing gene expression in human placenta -

XX Claim 27; SEQ ID NO 33379; 654pp; English.

XX The present invention relates to single exon nucleic acid probes (SENPs;  
CC see A13315-A157546). The present sequence is a peptide encoded by one  
CC such probe. The probes are useful for producing a microarray for  
CC predicting, measuring and displaying gene expression in samples derived  
CC from human placenta. The probes are useful for antenatal diagnosis of  
CC human genetic disorders.

SQ Sequence 68 AA;

Query Match 31.2%; Score 5; DB 22; Length 68;

Best Local Similarity 100.0%; Pred. No. 2.7e+02;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 YGRDD 7  
|||

Db 40 YGRDD 44

Search completed: April 9, 2003, 13:12:31

Job time : 36.377 secs

**THIS PAGE BLANK (USPTO)**

















977 4 25.0 329 3 US-08-964-308-4 Sequence 4, Appli  
978 4 25.0 329 3 US-08-964-308-14 Sequence 14, Appli  
979 4 25.0 329 3 US-08-964-308-15 Sequence 15, Appli  
980 4 25.0 329 3 US-08-964-313-4 Sequence 4, Appli  
981 4 25.0 329 3 US-08-964-313-14 Sequence 14, Appli  
982 4 25.0 329 3 US-08-964-313-15 Sequence 15, Appli  
983 4 25.0 329 4 US-08-860-255A-1 Sequence 1, Appli  
984 4 25.0 329 4 US-09-069-138-4 Sequence 4, Appli  
985 4 25.0 329 4 US-09-069-138-14 Sequence 14, Appli  
986 4 25.0 329 4 US-09-069-138-15 Sequence 15, Appli  
987 4 25.0 329 4 US-08-915-095A-2 Sequence 2, Appli  
988 4 25.0 329 4 US-08-915-095A-7 Sequence 7, Appli  
989 4 25.0 329 4 US-08-798-096-2 Sequence 2, Appli  
990 4 25.0 329 4 US-08-798-096-7 Sequence 7, Appli  
991 4 25.0 329 4 US-08-684-932A-36 Sequence 36, Appli  
992 4 25.0 329 4 US-08-798-095A-2 Sequence 2, Appli  
993 4 25.0 329 4 US-08-798-095A-7 Sequence 7, Appli  
994 4 25.0 329 5 PCT-US94-04781-2 Sequence 2, Appli  
995 4 25.0 329 5 PCT-US95-13820-2 Sequence 2, Appli  
996 4 25.0 329 5 PCT-US95-13820-3 Sequence 3, Appli  
997 4 25.0 331 1 US-08-136-993-1 Sequence 1, Appli  
998 4 25.0 331 1 US-08-208-007A-8 Sequence 8, Appli  
999 4 25.0 331 1 US-08-330-121B-4 Sequence 4, Appli  
1000 4 25.0 331 3 US-09-109-063-1 Sequence 1, Appli

## ALIGNMENTS

RESULT 1  
US-09-142-078-66  
; Sequence 66, Application US/09142078  
; Patent No. 6172041  
; GENERAL INFORMATION:  
; APPLICANT: McCabe, R. Tyler  
; APPLICANT: Zhou, Li-Ming  
; APPLICANT: Layer, Richard T.  
; TITLE OF INVENTION: Use of Conantokins  
; NUMBER OF SEQUENCES: 71  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Rothwell, Figg, Ernst & Kurz, P.C.  
; STREET: 555 Thirteenth Street, N.W., Suite 701-E  
; CITY: Washington  
; STATE: D.C.  
; COUNTRY: USA  
; ZIP: 20004  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentin Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/142,078  
; FILING DATE: 10-FEB-1999  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: WO US97/12652  
; FILING DATE: 21-JUL-1997  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/762,377  
; FILING DATE: 06-DEC-1996  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/684,750  
; FILING DATE: 22-JUL-1996  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Ihnen, Jeffrey L.  
; REGISTRATION NUMBER: 28,957  
; REFERENCE/DOCKET NUMBER: 2314-135-A  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 202-783-6040  
; TELEFAX: 202-783-6031  
; INFORMATION FOR SEQ ID NO: 66:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 95 amino acids

TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-09-142-078-66

Query Match 37.5%; Score 6; DB 4; Length 95;  
Best Local Similarity 100.0%; Pred. No. 8.4;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 6 DDNRD 11  
DB 51 DDNRD 56

RESULT 2  
US-09-357-141-66  
; Sequence 66, Application US/09357141  
; Patent No. 6277825  
; GENERAL INFORMATION:  
; APPLICANT: Oliveira, Baldomero M.  
; APPLICANT: McIntosh, J. Michael  
; APPLICANT: McCabe, R. Tyler  
; APPLICANT: Layer, Richard T.  
; APPLICANT: Zhou, Li-Ming  
; TITLE OF INVENTION: Use of Conantokins for Treating Pain  
; FILE REFERENCE: 2314-171  
; CURRENT APPLICATION NUMBER: US/09/357,141  
; CURRENT FILING DATE: 1999-07-20  
; PRIOR APPLICATION NUMBER: US 09/283,277  
; PRIOR FILING DATE: 1999-04-01  
; PRIOR APPLICATION NUMBER: US 09/142,078  
; PRIOR FILING DATE: 1999-02-10  
; PRIOR APPLICATION NUMBER: WO US97/12652  
; PRIOR FILING DATE: 1997-07-21  
; PRIOR APPLICATION NUMBER: US 08/762,377  
; PRIOR FILING DATE: 1996-12-06  
; PRIOR APPLICATION NUMBER: US 08/684,750  
; PRIOR FILING DATE: 1996-07-22  
; NUMBER OF SEQ ID NOS: 71  
; SOFTWARE: Patentin Ver. 2.0  
; SEQ ID NO 66  
; LENGTH: 95  
; TYPE: PRT  
; ORGANISM: Conus quercinus  
US-09-357-141-66

Query Match 37.5%; Score 6; DB 4; Length 95;  
Best Local Similarity 100.0%; Pred. No. 8.4;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 6 DDNRD 11  
DB 51 DDNRD 56

RESULT 3  
US-09-533-889-66  
; Sequence 66, Application US/09533889  
; Patent No. 6399574  
; GENERAL INFORMATION:  
; APPLICANT: McCabe, R. Tyler  
; APPLICANT: Zhou, Li-Ming  
; APPLICANT: Layer, Richard T.  
; APPLICANT: Oliveira, Baldomero M.  
; APPLICANT: McIntosh, J. Michael  
; TITLE OF INVENTION: Use of Conantokins  
; NUMBER OF SEQUENCES: 71  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Rothwell, Figg, Ernst & Kurz, P.C.  
; STREET: 555 Thirteenth Street, N.W., Suite 701-E  
; CITY: Washington  
; STATE: D.C.  
; COUNTRY: USA

```
ZIP: 20004
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/533,889
; FILING DATE: 22 MAR-2000
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 09/142,078
; FILING DATE: 10-FEB-1999
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: WO US97/12652
; FILING DATE: 21-JUL-1997
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/762,377
; FILING DATE: 06-DEC-1996
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/684,750
; FILING DATE: 22-JUL-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Ihnen, Jeffrey L.
; REGISTRATION NUMBER: 28,957
; REFERENCE/DOCKET NUMBER: 2314-168.A
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-783-6040
; TELEFAX: 202-783-6031
; INFORMATION FOR SEQ ID NO: 66:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 95 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-09-533-889-66
```

```
Query Match 37.5%; Score 6; DB 4; Length 95;
Best Local Similarity 100.0%; Pred. No. 8.4;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 6 DDNRD 11
|||
Db 51 DDNRD 56
```

```
RESULT 4
US-09-512-342-20
; Sequence 20, Application US/09512342
; Patent No. 6388068
; GENERAL INFORMATION:
; APPLICANT: SATOH, SHINOBU
; APPLICANT: MASUDA, SUSUMU
; TITLE OF INVENTION: METHOD FOR PRODUCING FOREIGN POLYPEPTIDE IN PLANT
; FILE REFERENCE: 081356/0142
; CURRENT APPLICATION NUMBER: US/09/512,342
; CURRENT FILING DATE: 2000-02-24
; NUMBER OF SEQ ID NOS: 38
; SOFTWARE: Patentin Ver. 2.1
; SEQ ID NO 20
; LENGTH: 168
; TYPE: PRT
; ORGANISM: Cucumis sativus
US-09-512-342-20
```

```
Query Match 37.5%; Score 6; DB 4; Length 168;
Best Local Similarity 100.0%; Pred. No. 13;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 1 VGYGRD 6
|||
Db 53 VGYGRD 58
```

```
RESULT 5
PCT-US93-05701-16
; Sequence 16, Application PC/TUS9305701
; GENERAL INFORMATION:
; APPLICANT: Massachusetts Institute of Technology
; TITLE OF INVENTION: Cloning and Characterization of Cell Death Genes
; NUMBER OF SEQUENCES: 29
; CORRESPONDENCE ADDRESS:
; ADDRESS: Massachusetts Institute of Technology
; STREET: 77 Massachusetts Avenue
; CITY: Cambridge
; STATE: Massachusetts
; COUNTRY: U.S.A.
; ZIP: 02139
; COMPUTER READABLE FORM:
; MEDIUM TYPE: diskette
; COMPUTER:
; OPERATING SYSTEM:
; SOFTWARE:
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US93/05701
; FILING DATE: 19930614
; CLASSIFICATION:
; INFORMATION FOR SEQ ID NO: 16:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 12 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
PCT-US93-05701-16
```

```
Query Match 31.2%; Score 5; DB 5; Length 12;
Best Local Similarity 100.0%; Pred. No. 17;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 11 DPSRK 15
|||
Db 3 DPSRK 7
```

```
RESULT 6
US-09-152-060-113
; Sequence 113, Application US/09152060
; Patent No. 6448230
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: 28 Human Secreted Proteins
; FILE REFERENCE: P2003P1.US
; CURRENT APPLICATION NUMBER: US/09/152,060
; CURRENT FILING DATE: 1998-09-11
; EARLIER APPLICATION NUMBER: PCT/US98/04858
; EARLIER FILING DATE: 1998-03-12
; EARLIER APPLICATION NUMBER: 60/040,762
; EARLIER FILING DATE: 1997-03-14
; EARLIER APPLICATION NUMBER: 60/040,710
; EARLIER FILING DATE: 1997-03-14
; EARLIER APPLICATION NUMBER: 60/050,934
; EARLIER FILING DATE: 1997-05-30
; EARLIER APPLICATION NUMBER: 60/048,100
; EARLIER FILING DATE: 1997-05-30
; EARLIER APPLICATION NUMBER: 60/048,357
; EARLIER FILING DATE: 1997-05-30
; EARLIER APPLICATION NUMBER: 60/048,189
; EARLIER FILING DATE: 1997-05-30
; EARLIER APPLICATION NUMBER: 60/057,765
; EARLIER FILING DATE: 1997-09-05
; EARLIER APPLICATION NUMBER: 60/048,970
; EARLIER FILING DATE: 1997-06-06
; EARLIER APPLICATION NUMBER: 60/068,368
; EARLIER FILING DATE: 1997-12-19
; NUMBER OF SEQ ID NOS: 118
; SOFTWARE: Patentin Ver. 2.0
```

; SEQ ID NO 113  
; LENGTH: 18  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-152-060-113

Query Match 31.2%; Score 5; DB 4; Length 18;  
Best Local Similarity 100.0%; Pred. No. 24;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2 GYGRD 6  
Db 4 GYGRD 8

RESULT 7  
PCT-US94-01234-2  
; Sequence 2, Application PC/TUS9401234  
; GENERAL INFORMATION:  
; APPLICANT:  
; TITLE OF INVENTION: METHODS FOR PRODUCING POLYPEPTIDE  
; TITLE OF INVENTION: BINDING SITES  
; NUMBER OF SEQUENCES: 76  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: PCT/US94/01234  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/084,542  
; FILING DATE: 28-JUN-1993  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/012,566  
; FILING DATE: 02-FEB-1993  
; INFORMATION FOR SEQ ID NO: 2:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 18 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; FRAGMENT TYPE: internal  
PCT-US94-01234-2

Query Match 31.2%; Score 5; DB 5; Length 18;  
Best Local Similarity 100.0%; Pred. No. 24;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 VGYGR 5  
Db 1 VGYGR 5

RESULT 8  
US-09-227-357-493  
; Sequence 493, Application US/09227357  
; Patent No. 6342581  
; GENERAL INFORMATION:  
; APPLICANT: Fischer et al.  
; TITLE OF INVENTION: 123 Human Secreted Proteins  
; FILE REFERENCE: P2010P1  
; CURRENT APPLICATION NUMBER: US/09/227,357  
; CURRENT FILING DATE: 1999-01-08  
; EARLIER APPLICATION NUMBER: PCT/US98/13684  
; EARLIER FILING DATE: 1998-07-07  
; EARLIER APPLICATION NUMBER: 60/051,926  
; EARLIER FILING DATE: 1997-07-08  
; EARLIER APPLICATION NUMBER: 60/052,793  
; EARLIER FILING DATE: 1997-07-08  
; EARLIER APPLICATION NUMBER: 60/051,925  
; EARLIER FILING DATE: 1997-07-08  
; EARLIER APPLICATION NUMBER: 60/051,929

; EARLIER FILING DATE: 1997-07-08  
; EARLIER APPLICATION NUMBER: 60/052,803  
; EARLIER FILING DATE: 1997-07-08  
; EARLIER APPLICATION NUMBER: 60/052,732  
; EARLIER FILING DATE: 1997-07-08  
; EARLIER APPLICATION NUMBER: 60/051,931  
; EARLIER FILING DATE: 1997-07-08  
; EARLIER APPLICATION NUMBER: 60/051,932  
; EARLIER FILING DATE: 1997-07-08  
; EARLIER APPLICATION NUMBER: 60/051,916  
; EARLIER FILING DATE: 1997-07-08  
; EARLIER APPLICATION NUMBER: 60/051,930  
; EARLIER FILING DATE: 1997-07-08  
; EARLIER APPLICATION NUMBER: 60/051,918  
; EARLIER FILING DATE: 1997-07-08  
; EARLIER APPLICATION NUMBER: 60/051,920  
; EARLIER FILING DATE: 1997-07-08  
; EARLIER APPLICATION NUMBER: 60/052,733  
; EARLIER FILING DATE: 1997-07-08  
; EARLIER APPLICATION NUMBER: 60/052,795  
; EARLIER FILING DATE: 1997-07-08  
; EARLIER APPLICATION NUMBER: 60/051,919  
; EARLIER FILING DATE: 1997-07-08  
; EARLIER APPLICATION NUMBER: 60/051,928  
; EARLIER FILING DATE: 1997-07-08  
; EARLIER APPLICATION NUMBER: 60/055,722  
; EARLIER FILING DATE: 1997-08-18  
; EARLIER APPLICATION NUMBER: 60/055,723  
; EARLIER FILING DATE: 1997-08-18  
; EARLIER APPLICATION NUMBER: 60/055,948  
; EARLIER FILING DATE: 1997-08-18  
; EARLIER APPLICATION NUMBER: 60/055,949  
; EARLIER FILING DATE: 1997-08-18  
; EARLIER APPLICATION NUMBER: 60/055,953  
; EARLIER FILING DATE: 1997-08-18  
; EARLIER APPLICATION NUMBER: 60/055,950  
; EARLIER FILING DATE: 1997-08-18  
; EARLIER APPLICATION NUMBER: 60/055,947  
; EARLIER FILING DATE: 1997-08-18  
; EARLIER APPLICATION NUMBER: 60/055,964  
; EARLIER FILING DATE: 1997-08-18  
; EARLIER APPLICATION NUMBER: 60/056,360  
; EARLIER FILING DATE: 1997-08-18  
; EARLIER APPLICATION NUMBER: 60/055,684  
; EARLIER FILING DATE: 1997-08-18  
; EARLIER APPLICATION NUMBER: 60/055,984  
; EARLIER FILING DATE: 1997-08-18  
; EARLIER APPLICATION NUMBER: 60/055,954  
; EARLIER FILING DATE: 1997-08-18  
; EARLIER APPLICATION NUMBER: 60/058,785  
; EARLIER FILING DATE: 1997-09-12  
; EARLIER APPLICATION NUMBER: 60/058,664  
; EARLIER FILING DATE: 1997-09-12  
; EARLIER APPLICATION NUMBER: 60/058,660  
; EARLIER FILING DATE: 1997-09-12  
; EARLIER APPLICATION NUMBER: 60/058,661  
; EARLIER FILING DATE: 1997-09-12  
; NUMBER OF SEQ ID NOS: 672  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 493  
; LENGTH: 24  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-227-357-493

Query Match 31.2%; Score 5; DB 4; Length 24;  
Best Local Similarity 100.0%; Pred. No. 30;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 10 RDPSSR 14  
Db 18 RDPSSR 22

RESULT 9  
US-09-065-383-30  
; Sequence 30, Application US/09065383  
; Patent No. 6391543  
; GENERAL INFORMATION:  
; APPLICANT: BILLING-MEDEL, PATRICIA  
; APPLICANT: COHEN, MAURICE  
; APPLICANT: COLPITTS, TRACEY L.  
; APPLICANT: FRIEDMAN, PAULA N.  
; APPLICANT: GORDON, JULIAN  
; APPLICANT: GRANDOS, EDWARD N.  
; APPLICANT: HODGES, STEVEN C.  
; APPLICANT: KLAS, MICHAEL R.  
; APPLICANT: KRATOCHVIL, JON D.  
; APPLICANT: ROBERTS-RAPP, LISA  
; APPLICANT: RUSSELL, JOHN C.  
; APPLICANT: STROUPE, STEPHEN D.  
; TITLE OF INVENTION: REAGENTS AND METHODS USEFUL  
; IN THE TITRATION OF FOR DETECTING DISEASES OF THE PROSTATE  
; NUMBER OF SEQUENCES: 33  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Abbott Laboratories  
; STREET: 100 Abbott Park Road  
; CITY: Abbott Park  
; STATE: IL  
; COUNTRY: USA  
; ZIP: 60064-3500  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette  
; COMPUTER: IBM Compatible  
; OPERATING SYSTEM: DOS  
; SOFTWARE: FastSeq for Windows Version 2.0  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/065.383  
; FILING DATE:  
; CLASSIFICATION:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 08/842.385  
; FILING DATE: 23-APR-1997  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Becker, Cheryl L.  
; REGISTRATION NUMBER: 35,441  
; REFERENCE/DOCKET NUMBER: 6084.US.P1  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 847/935-1729  
; TELEFAX: 847/938-2623  
; TELEX:  
; INFORMATION FOR SEQ ID NO: 30:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 40 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: No. 6391543e  
; US-09-065-383-30

Query Match 31.2%; Score 5; DB 4; Length 40;  
Best Local Similarity 100.0%; Pred. No. 46;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 11 DPSRK 15  
|||  
Db 8 DPSRK 12

RESULT 10  
US-08-659-251-31  
; Sequence 31, Application US/08659251  
; Patent No. 5883081  
; GENERAL INFORMATION:  
; APPLICANT: Kraus, Guenter  
; APPLICANT: Wong-Staal, Flossie

; APPLICANT: Talbot, Randy  
; APPLICANT: Poeschla, Eric  
; TITLE OF INVENTION: Isolation of No. 5883081e1 HIV-2 Proviruses  
; NUMBER OF SEQUENCES: 50  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Townsend and Townsend and Crew LLP  
; STREET: Two Embarcadero Center, Eighth Floor  
; CITY: San Francisco  
; STATE: California  
; COUNTRY: USA  
; ZIP: 94111-3834  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/659.251  
; FILING DATE: No. 5883081 yet assigned  
; CLASSIFICATION: 514  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 60/001,441  
; FILING DATE: 26-JUL-1995  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Garrett-Wackowski, Eugenia  
; REGISTRATION NUMBER: 37,330  
; REFERENCE/DOCKET NUMBER: 02307E-056410US  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (415) 576-0200  
; TELEFAX: (415) 576-0300  
; INFORMATION FOR SEQ ID NO: 31:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 68 amino acids  
; TYPE: amino acid  
; STRANDEDNESS:  
; TOPOLOGY: linear  
; MOLECULE TYPE: peptide  
; FEATURE:  
; NAME/KEY: Peptide  
; LOCATION: 1..68  
; OTHER INFORMATION: /note= "HIV-2H0 rev protein amino acids  
; OTHER INFORMATION: homologous to HIV-2KR rev protein amino  
; OTHER INFORMATION: acid residues from positions 58 to 160"  
; US-08-659-251-31

Query Match 31.2%; Score 5; DB 2; Length 68;  
Best Local Similarity 100.0%; Pred. No. 71;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2 GYGRD 6  
|||  
Db 34 GYGRD 38

RESULT 11  
US-09-256-490-31  
; Sequence 31, Application US/09256490  
; Patent No. 6235881  
; GENERAL INFORMATION:  
; APPLICANT: Kraus, Guenter  
; APPLICANT: Wong-Staal, Flossie  
; APPLICANT: Talbot, Randy  
; APPLICANT: Poeschla, Eric  
; TITLE OF INVENTION: Isolation of No. 6235881e1 HIV-2 Proviruses  
; NUMBER OF SEQUENCES: 50  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Townsend and Townsend and Crew LLP  
; STREET: Two Embarcadero Center, Eighth Floor  
; CITY: San Francisco  
; STATE: California  
; COUNTRY: USA  
; ZIP: 94111-3834  
; COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/256,490  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/659,251  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: Garrett-Wackowski, Eugenia  
REGISTRATION NUMBER: 37,330  
REFERENCE/DOCKET NUMBER: 02307E-056410US  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (415) 576-0200  
TELEFAX: (415) 576-0300  
INFORMATION FOR SEQ ID NO: 31:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 68 amino acids  
TYPE: amino acid  
STRANDEDNESS:  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
FEATURE:  
NAME/KEY: Peptide  
LOCATION: 1..68  
OTHER INFORMATION: /note= "HIV-2EHO rev protein amino acids homologous to HIV-2KR rev protein amino  
OTHER INFORMATION: acid residues from positions 58 to 160"  
US-09-256-490-31

Query Match 31.2%; Score 5; DB 4; Length 68;  
Best Local Similarity 100.0%; Pred. No. 71;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 2 GYGRD 6  
|||  
Db 34 GYGRD 38

RESULT 12  
PCT-US96-11445-31  
Sequence 31, Application PC/TUS9611445  
GENERAL INFORMATION:  
APPLICANT: The Regents of the University of California  
TITLE OF INVENTION: Isolation of Novel HIV-2 Proviruses  
NUMBER OF SEQUENCES: 50  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Robbins, Berliner & Carson  
STREET: 201 N. Figueroa Street, 5th Floor  
CITY: Los Angeles  
STATE: California  
COUNTRY: USA  
ZIP: 90012-2628  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: PCT/US96/11445  
FILING DATE:  
CLASSIFICATION:  
ATTORNEY/AGENT INFORMATION:  
NAME: Berliner, Robert  
REGISTRATION NUMBER: 20,121  
REFERENCE/DOCKET NUMBER: 5555-399C1  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (213) 977-1001  
TELEFAX: (213) 977-1003  
INFORMATION FOR SEQ ID NO: 31:

SEQUENCE CHARACTERISTICS:  
LENGTH: 68 amino acids  
TYPE: amino acid  
STRANDEDNESS:  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
FEATURE:  
NAME/KEY: Peptide  
LOCATION: 1..68  
OTHER INFORMATION: /note= "HIV-2EHO rev protein amino acids homologous to HIV-2KR rev protein amino  
OTHER INFORMATION: acid residues from positions 58 to 160"  
PCT-US96-11445-31

Query Match 31.2%; Score 5; DB 5; Length 68;  
Best Local Similarity 100.0%; Pred. No. 71;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 2 GYGRD 6  
|||  
Db 34 GYGRD 38

RESULT 13  
US-08-422-613-4  
Sequence 4, Application US/08422613  
Patent No. 5695389  
GENERAL INFORMATION:  
APPLICANT: Louvard, Daniel  
APPLICANT: Dudonet, Brigitte  
APPLICANT: Robine, Sylvie  
APPLICANT: Arpin, Monique  
APPLICANT: Pingault, Eric  
APPLICANT: Garcia, Alphonse  
TITLE OF INVENTION: Agents for the In Vitro Diagnosis of  
TITLE OF INVENTION: Malignant Cells Originating in the Digestive Tract  
NUMBER OF SEQUENCES: 4  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Finnegan, Henderson, Farbow, Garrett &  
ADDRESS: Dunner  
STREET: 1300 I Street, N.W., Suite 700  
CITY: Washington  
STATE: D.C.  
COUNTRY: USA  
ZIP: 20005-3315  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/422,613  
FILING DATE:  
CLASSIFICATION: 530  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/989,696  
FILING DATE: 14-DEC-1992  
APPLICATION NUMBER: US 07/662,992  
FILING DATE: 28-FEB-1991  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: WO PCT/FR86.00150  
FILING DATE: 30-APR-1986  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: FR 85.06707  
FILING DATE: 02-MAY-1985  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: FR 85.16820  
FILING DATE: 13-NOV-1985  
ATTORNEY/AGENT INFORMATION:  
NAME: Potter, Jane E.  
REGISTRATION NUMBER: 33,332  
REFERENCE/DOCKET NUMBER: 02356-0006-04000  
TELECOMMUNICATION INFORMATION:



TELEPHONE: 202-408-4000  
TELEFAX: 202-408-4400  
INFORMATION FOR SEQ ID NO: 4:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 75 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
US-08-422-613-4

Query Match 31.2%; Score 5; DB 1; Length 75;  
Best Local Similarity 100.0%; Pred. No. 77;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 11 PSRK 15  
Db 34 PSRK 38

RESULT 14  
US-08-202-389-26  
Sequence 26, Application US/08202389  
Patent No. 5536636  
GENERAL INFORMATION:  
APPLICANT: Freeman Jr., Robert M.  
APPLICANT: Plutsky, Jorge  
APPLICANT: Neel, Benjamin G.  
APPLICANT: Rosenberg, Robert D.  
TITLE OF INVENTION: IDENTIFICATION OF NOVEL TYROSINE  
NUMBER OF SEQUENCES: 54  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Hamilton, Brook, Smith & Reynolds, P.C.  
STREET: Two Millicia Drive  
CITY: Lexington  
STATE: MA  
COUNTRY: USA  
ZIP: 02173  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/202.389  
FILING DATE: 28-FEB-1994  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/983,926  
FILING DATE: 01-DEC-1992  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/829,141  
FILING DATE: 31-JAN-1992  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/721,112  
FILING DATE: 26-JUN-1991  
ATTORNEY/AGENT INFORMATION:  
NAME: Granahan, Patricia  
REGISTRATION NUMBER: 32,227  
REFERENCE/DOCKET NUMBER: BIH92-05WA  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (617) 861-6240  
TELEFAX: (617) 861-9540  
INFORMATION FOR SEQ ID NO: 26:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 100 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-202-389-26

Query Match 31.2%; Score 5; DB 1; Length 100;  
Best Local Similarity 100.0%; Pred. No. 97;

Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Qy 12 PSRK 16  
Db 28 PSRK 32

RESULT 15  
US-08-422-613-2  
Sequence 2, Application US/08422613  
Patent No. 5695389  
GENERAL INFORMATION:  
APPLICANT: Louvard, Daniel  
APPLICANT: Dudouet, Brigitte  
APPLICANT: Robine, Sylvie  
APPLICANT: Arpin, Monique  
APPLICANT: Pringault, Eric  
APPLICANT: Garcia, Alphonse  
TITLE OF INVENTION: Agents for the In Vitro Diagnosis of  
TITLE OF INVENTION: Malignant Cells Originating in the Digestive Tract  
NUMBER OF SEQUENCES: 4  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Finnegan, Henderson, Farabow, Garrett &  
ADDRESS: Dunner  
STREET: 1300 I Street, N.W., Suite 700  
CITY: Washington  
STATE: D.C.  
COUNTRY: USA  
ZIP: 20005-3315  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/422.613  
FILING DATE:  
CLASSIFICATION: 530  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/989,696  
FILING DATE: 14-DEC-1992  
APPLICATION NUMBER: US 07/662,992  
FILING DATE: 28-FEB-1991  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: WO PCT/FR86.00150  
FILING DATE: 30-APR-1986  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: FR 85.06707  
FILING DATE: 02-MAY-1985  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: FR 85.16820  
FILING DATE: 13-NOV-1985  
ATTORNEY/AGENT INFORMATION:  
NAME: Potler, Jane E.  
REGISTRATION NUMBER: 33,332  
REFERENCE/DOCKET NUMBER: 02356-0006-04000  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 202-408-4000  
TELEFAX: 202-408-4400  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 110 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
US-08-422-613-2

Query Match 31.2%; Score 5; DB 1; Length 110;  
Best Local Similarity 100.0%; Pred. No. 11e+02;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Qy 11 PSRK 15  
Db 34 PSRK 38

Db 68 DPSRK 72

## RESULT 16

US-09-188-930-179  
; Sequence 179, Application US/09188930A  
; Patent No. 6150502  
; GENERAL INFORMATION:  
; APPLICANT: Watson, James D.  
; APPLICANT: Strachan, Lorna  
; APPLICANT: Sleeman, Matthew  
; APPLICANT: Onrust, Rene  
; APPLICANT: Murison, James Greg  
; TITLE OF INVENTION: Compositions Isolated From Skin Cells  
; TITLE OF INVENTION: and Methods For Their Use  
; FILE REFERENCE: 11000.1011c1  
; CURRENT APPLICATION NUMBER: US/09/188.930A  
; CURRENT FILING DATE: 1998-11-09  
; NUMBER OF SEQ ID NOS: 348  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 179  
; LENGTH: 123  
; TYPE: PRT  
; ORGANISM: mouse  
US-09-188-930-179

Query Match 31.2%; Score 5; DB 4; Length 123;

Best Local Similarity 100.0%; Pred. No. 1.2e+02;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 12 PSRKN 16

Db 113 PSRKN 117

## RESULT 17

US-07-644-372-2  
; Sequence 2, Application US/07644372  
; Patent No. 5416009  
; GENERAL INFORMATION:  
; APPLICANT: Lazzeri, Mario E.  
; APPLICANT: Nutman, Thomas B.  
; APPLICANT: Weiss, Niklaus  
; TITLE OF INVENTION: A DNA SEGMENT ENCODING A SPECIFIC  
; TITLE OF INVENTION: IMMUNODIAGNOSTIC ANTIGEN  
; NUMBER OF SEQUENCES: 2  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: CUSHMAN, DARBY & CUSHMAN  
; STREET: 1615 L. Street, N.W.  
; CITY: Washington  
; STATE: D.C.  
; COUNTRY: USA  
; ZIP: 20036  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/07/644.372  
; FILING DATE: 19910123  
; CLASSIFICATION: 435  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (202)861-3000  
; TELEFAX: (202)822-0944  
; INFORMATION FOR SEQ ID NO: 2:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 152 amino acids  
; TYPE: AMINO ACID  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-07-644-372-2

Query Match 31.2%; Score 5; DB 1; Length 152;  
Best Local Similarity 100.0%; Pred. No. 1.4e+02;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 12 PSRKN 16

Db 102 PSRKN 106

## RESULT 18

US-09-208-718-6  
; Sequence 6, Application US/09208718  
; Patent No. 6063767  
; GENERAL INFORMATION:  
; APPLICANT: Lal, Preeti  
; APPLICANT: Hillman, Jennifer  
; APPLICANT: Corley, Neil  
; APPLICANT: Shah, Purvi  
; TITLE OF INVENTION: HUMAN PHOSPHOLIPID BINDING PROTEINS  
; NUMBER OF SEQUENCES: 6  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Incyte Pharmaceuticals, Inc.  
; STREET: 3174 Porter Dr.  
; CITY: Palo Alto  
; STATE: CA  
; COUNTRY: USA  
; ZIP: 94304  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette  
; COMPUTER: IBM Compatible  
; OPERATING SYSTEM: DOS  
; SOFTWARE: FastSeq for Windows Version 2.0  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/208.718  
; FILING DATE:  
; CLASSIFICATION:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 08/958,820  
; FILING DATE:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Billings, Lucy J.  
; REGISTRATION NUMBER: 36,749  
; REFERENCE/DOCKET NUMBER: PF-0379 US  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 650-855-0555  
; TELEFAX: 650-845-4166  
; INFORMATION FOR SEQ ID NO: 6:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 152 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; IMMEDIATE SOURCE:  
; LIBRARY: Genbank  
; CLONE: 1143527  
US-09-208-718-6

Query Match 31.2%; Score 5; DB 3; Length 152;

Best Local Similarity 100.0%; Pred. No. 1.4e+02;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 12 PSRKN 16

Db 57 PSRKN 61

## RESULT 19

US-09-562-737-115  
; Sequence 115, Application US/09562737  
; Patent No. 6428967  
; GENERAL INFORMATION:  
; APPLICANT: Hertz, Joachim  
; APPLICANT: Gotthardt, Michael

```

; TITLE OF INVENTION: LDL Receptor Signaling Pathways
; FILE REFERENCE: UTSW0708
; CURRENT APPLICATION NUMBER: US/09/562,737
; CURRENT FILING DATE: 2000-05-01
; NUMBER OF SEQ ID NOS: 132
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 115
; LENGTH: 185
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Synthetic
US-09-562-737-115

Query Match
Best Local Similarity 31.2%; Score 5; DB 4; Length 185;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 4 GRDDN 8
    |||||
Db 45 GRDDN 49

RESULT 20
; Sequence 145, Application US/09325932A
; Patent No. 6451604
; GENERAL INFORMATION:
; APPLICANT: Flihn, Barry
; APPLICANT: Laaham, Annette
; TITLE OF INVENTION: Compositions affecting programmed cell
; FILE REFERENCE: 1022
; CURRENT APPLICATION NUMBER: US/09/325,932A
; CURRENT FILING DATE: 1999-06-04
; NUMBER OF SEQ ID NOS: 206
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 145
; LENGTH: 204
; TYPE: PRT
; ORGANISM: Eucalyptus grandis
US-09-325-932A-145

Query Match
Best Local Similarity 31.2%; Score 5; DB 4; Length 204;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 VGYGR 5
    |||||
Db 128 VGYGR 132

RESULT 21
US-08-622-354-5
; Sequence 5, Application US/08622354
; Patent No. 5827518
; GENERAL INFORMATION:
; APPLICANT: WEBB, Bruce A.
; APPLICANT: CUI, Liwang
; TITLE OF INVENTION: VIRAL AND INSECT GENES THAT INHIBIT THE
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: LOWE, PRICE, LEBLANC & BECKER
; STREET: 99 Canal Center Plaza, Suite 300
; CITY: Alexandria
; STATE: VA
; COUNTRY: US
; ZIP: 22314
; COMPUTER READABLE FORM:
; MEDIUM TYPE: floppy disk
; COMPUTER: IBM PC compatible
```

```

; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/622,354
; FILING DATE: 27-MAR-1996
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Price, Robert L.
; REGISTRATION NUMBER: 22,685
; REFERENCE/DOCKET NUMBER: 434-061
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 684-1111
; TELEFAX: (703) 684-1124
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 231 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-622-354-5

Query Match
Best Local Similarity 31.2%; Score 5; DB 2; Length 231;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 12 PSRKN 16
    |||||
Db 145 PSRKN 149

RESULT 22
US-09-227-357-488
; Sequence 488, Application US/09227357
; Patent No. 6342581
; GENERAL INFORMATION:
; APPLICANT: Fischer et al.
; TITLE OF INVENTION: 123 Human Secreted Proteins
; FILE REFERENCE: P2010P1
; CURRENT APPLICATION NUMBER: US/09/227,357
; CURRENT FILING DATE: 1999-01-08
; EARLIER APPLICATION NUMBER: PCT/US98/13684
; EARLIER FILING DATE: 1998-07-07
; EARLIER APPLICATION NUMBER: 60/051,926
; EARLIER FILING DATE: 1997-07-08
; EARLIER APPLICATION NUMBER: 60/052,793
; EARLIER FILING DATE: 1997-07-08
; EARLIER APPLICATION NUMBER: 60/051,925
; EARLIER FILING DATE: 1997-07-08
; EARLIER APPLICATION NUMBER: 60/051,929
; EARLIER FILING DATE: 1997-07-08
; EARLIER APPLICATION NUMBER: 60/052,803
; EARLIER FILING DATE: 1997-07-08
; EARLIER APPLICATION NUMBER: 60/052,732
; EARLIER FILING DATE: 1997-07-08
; EARLIER APPLICATION NUMBER: 60/051,931
; EARLIER FILING DATE: 1997-07-08
; EARLIER APPLICATION NUMBER: 60/051,932
; EARLIER FILING DATE: 1997-07-08
; EARLIER APPLICATION NUMBER: 60/051,916
; EARLIER FILING DATE: 1997-07-08
; EARLIER APPLICATION NUMBER: 60/051,930
; EARLIER FILING DATE: 1997-07-08
; EARLIER APPLICATION NUMBER: 60/051,918
; EARLIER FILING DATE: 1997-07-08
; EARLIER APPLICATION NUMBER: 60/051,920
; EARLIER FILING DATE: 1997-07-08
; EARLIER APPLICATION NUMBER: 60/052,733
; EARLIER FILING DATE: 1997-07-08
; EARLIER APPLICATION NUMBER: 60/052,795
; EARLIER FILING DATE: 1997-07-08
; EARLIER APPLICATION NUMBER: 60/051,919
; EARLIER FILING DATE: 1997-07-08
; EARLIER APPLICATION NUMBER: 60/051,928
```

```
; EARLIER FILING DATE: 1997-07-08
; EARLIER APPLICATION NUMBER: 60/055,722
; EARLIER FILING DATE: 1997-08-18
; EARLIER APPLICATION NUMBER: 60/055,723
; EARLIER FILING DATE: 1997-08-18
; EARLIER APPLICATION NUMBER: 60/055,948
; EARLIER FILING DATE: 1997-08-18
; EARLIER APPLICATION NUMBER: 60/055,949
; EARLIER FILING DATE: 1997-08-18
; EARLIER APPLICATION NUMBER: 60/055,953
; EARLIER FILING DATE: 1997-08-18
; EARLIER APPLICATION NUMBER: 60/055,950
; EARLIER FILING DATE: 1997-08-18
; EARLIER APPLICATION NUMBER: 60/055,947
; EARLIER FILING DATE: 1997-08-18
; EARLIER APPLICATION NUMBER: 60/055,964
; EARLIER FILING DATE: 1997-08-18
; EARLIER APPLICATION NUMBER: 60/056,360
; EARLIER FILING DATE: 1997-08-18
; EARLIER APPLICATION NUMBER: 60/055,684
; EARLIER FILING DATE: 1997-08-18
; EARLIER APPLICATION NUMBER: 60/055,984
; EARLIER FILING DATE: 1997-08-18
; EARLIER APPLICATION NUMBER: 60/055,954
; EARLIER FILING DATE: 1997-08-18
; EARLIER APPLICATION NUMBER: 60/058,785
; EARLIER FILING DATE: 1997-09-12
; EARLIER APPLICATION NUMBER: 60/058,664
; EARLIER FILING DATE: 1997-09-12
; EARLIER APPLICATION NUMBER: 60/058,660
; EARLIER FILING DATE: 1997-09-12
; EARLIER APPLICATION NUMBER: 60/058,661
; EARLIER FILING DATE: 1997-09-12
; NUMBER OF SEQ ID NOS: 672
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 488
; LENGTH: 255
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-227-357-488
```

```
Query Match      31.2%; Score 5; DB 4; Length 255;
Best Local Similarity 100.0%; Pred. No. 2.1e+02;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
Qy      10 RDPSS 14
      |||||
Db      245 RDPSS 249
```

```
RESULT 23
US-08-675-921B-9
; Sequence 9, Application US/08675921B
; Patent No. 5863728
; GENERAL INFORMATION:
; APPLICANT: John Siu-Cheong Ho, John T. Loh, Melvin
; APPLICANT: Schindler and John L. Wang
; TITLE OF INVENTION: DNA Encoding Carbohydrate
; TITLE OF INVENTION: Binding Protein and Biological
; TITLE OF INVENTION: Materials Derived Therefrom
; NUMBER OF SEQUENCES: 9
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Ian C. McLeod
; STREET: 2190 Commons Parkway
; CITY: Okemos
; STATE: Michigan
; COUNTRY: USA
; ZIP: 48864
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette 5.25 inch, 360
; MEDIUM TYPE: kb storage
; COMPUTER: IBM compatible
; OPERATING SYSTEM: MS-DOS (version 3.3)
```

```
; SOFTWARE: Wordperfect 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/675,921B
; FILING DATE: 06/05/96
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Ian C. McLeod
; REGISTRATION NUMBER: 20,931
; REFERENCE/DOCKET NUMBER: MSU 4.1-265
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (517) 347-4100
; TELEFAX: (517) 347-4103
; TELEX: No. 5863728e
; INFORMATION FOR SEQ ID NO: 9:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 260
; TYPE: Amino Acid
; STRANDEDNESS: Double
; TOPOLOGY: Linear
; MOLECULE TYPE:
; DESCRIPTION: Peptide
; HYPOTHEICAL: NO
; ANTI-SENSE: NO
; FRAGMENT TYPE:
; ORIGINAL SOURCE:
; ORGANISM: Bradyrhizobium japonicum
; STRAIN:
; INDIVIDUAL ISOLATE:
; DEVELOPMENTAL STAGE: N/A
; HAPLOTYPE: N/A
; TISSUE TYPE: N/A
; CELL TYPE: N/A
; ORGANELLE: N/A
; IMMEDIATE SOURCE:
; POSITION IN GENOME: N/A
; FEATURES:
; NAME/KEY: DNA encoding carbohydrate
; NAME/KEY: binding protein
; LOCATION: chromosome DNA
; IDENTIFICATION METHOD: sequencing
; OTHER INFORMATION: promotes binding to
; OTHER INFORMATION: soybean
; PUBLICATION INFORMATION: N/A
US-08-675-921B-9
```

```
Query Match      31.2%; Score 5; DB 2; Length 260;
Best Local Similarity 100.0%; Pred. No. 2.1e+02;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
Qy      12 PRRKN 16
      |||||
Db      237 PRRKN 241
```

```
RESULT 24
US-08-461-731-2
; Sequence 2, Application US/08461731
; Patent No. 5716806
; GENERAL INFORMATION:
; APPLICANT: MEISSNER, ET AL.
; TITLE OF INVENTION: Human Inositol Monophosphatase H1
; NUMBER OF SEQUENCES: 9
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: CAPELLA, BYRNE, BAIN, GILFILLAN,
; ADDRESSEE: CECCHI, STEWART & OLSTEIN
; STREET: 6 BECKER FARM ROAD
; CITY: ROSELAND
; STATE: NEW JERSEY
; COUNTRY: USA
```

ZIP: 07068  
COMPUTER READABLE FORM:  
MEDIUM TYPE: 3.5 INCH DISKETTE  
COMPUTER: IBM PS/2  
OPERATING SYSTEM: MS-DOS  
SOFTWARE: WORD PERFECT 5.1  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/461,731  
FILING DATE: June 5, 1995  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: PCT/US94/10465  
FILING DATE: 16 SEP 94  
ATTORNEY/AGENT INFORMATION:  
NAME: FERRARO, GREGORY D.  
REGISTRATION NUMBER: 36,134  
REFERENCE/DOCKET NUMBER: 325800-316  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 201-994-1700  
TELEFAX: 201-994-1744  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 265 AMINO ACIDS  
TYPE: AMINO ACID  
STRANDEDNESS:  
TOPOLOGY: LINEAR  
MOLECULE TYPE: PROTEIN  
US-08-461-731-2

Query Match 31.2%; Score 5; DB 1; Length 265;  
Best Local Similarity 100.0%; Pred. No. 2.2e+02;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 3 YGRDD 7  
Db 259 YGRDD 263

RESULT 25  
US-09-002-072B-2  
Sequence 2, Application US/09002072B  
Patent No. 5953539  
GENERAL INFORMATION:  
APPLICANT: Meissner, Paul et al.  
TITLE OF INVENTION: Human Inositol Monophosphatase H1  
NUMBER OF SEQUENCES: 10  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Human Genome Science, Inc.  
STREET: 9410 Key West Avenue  
CITY: Rockville  
STATE: MD  
COUNTRY: USA  
ZIP: 20850  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/002,072B  
FILING DATE: 09-FEB-1998  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/461,731  
FILING DATE: 05-JUN-1995  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: WO PCT/US94/10465  
FILING DATE: 16-SEP-1994  
ATTORNEY/AGENT INFORMATION:  
NAME: Brookes, A. Anders  
REGISTRATION NUMBER: 36,373  
REFERENCE/DOCKET NUMBER: PF134D1  
TELECOMMUNICATION INFORMATION:

TELEPHONE: 301-309-8504  
TELEFAX: 301-309-8439  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 265 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-09-002-072B-2

Query Match 31.2%; Score 5; DB 2; Length 265;  
Best Local Similarity 100.0%; Pred. No. 2.2e+02;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 3 YGRDD 7  
Db 259 YGRDD 263

RESULT 26  
US-09-314-199-2  
Sequence 2, Application US/09314199  
Patent No. 6130051  
GENERAL INFORMATION:  
APPLICANT: MEISSNER, ET AL.  
TITLE OF INVENTION: Human Inositol Monophosphatase H1  
NUMBER OF SEQUENCES: 9  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: CARELLA, BYRNE, BAIN, GILFILLAN,  
ADDRESSEE: CECCHI, STEWART & OLSTEIN  
STREET: 6 BECKER FARM ROAD  
CITY: ROSELAND  
STATE: NEW JERSEY  
COUNTRY: USA  
ZIP: 07068

COMPUTER READABLE FORM:  
MEDIUM TYPE: 3.5 INCH DISKETTE  
COMPUTER: IBM PS/2  
OPERATING SYSTEM: MS-DOS  
SOFTWARE: WORD PERFECT 5.1  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/314,199  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 09/002,072  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: FERRARO, GREGORY D.  
REGISTRATION NUMBER: 36,134  
REFERENCE/DOCKET NUMBER: 325800-316  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 201-994-1700  
TELEFAX: 201-994-1744  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 265 AMINO ACIDS  
TYPE: AMINO ACID  
STRANDEDNESS:  
TOPOLOGY: LINEAR  
MOLECULE TYPE: PROTEIN  
US-09-314-199-2

Query Match 31.2%; Score 5; DB 4; Length 265;  
Best Local Similarity 100.0%; Pred. No. 2.2e+02;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 3 YGRDD 7  
Db 259 YGRDD 263

RESULT 27  
US-09-314-198-2  
; Sequence 2, Application US/09114198  
; Patent No. 640310  
; GENERAL INFORMATION:  
; APPLICANT: Meisner, Paul et al.  
; TITLE OF INVENTION: Human Inositol Monophosphatase H1  
; NUMBER OF SEQUENCES: 10  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Human Genome Science, Inc.  
; STREET: 9410 Key West Avenue  
; CITY: Rockville  
; STATE: MD  
; COUNTRY: USA  
; ZIP: 20850  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/314,198  
; FILING DATE:  
; CLASSIFICATION:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 09/002,072  
; FILING DATE:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: WO PCT/US94/10465  
; FILING DATE: 16-SEP-1994  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Brookes, A. Anders  
; REGISTRATION NUMBER: 36,373  
; REFERENCE/DOCKET NUMBER: PF134D1  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 301-309-8504  
; TELEFAX: 301-309-8439  
; INFORMATION FOR SEQ ID NO: 2:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 265 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; US-09-314-198-2  
Query Match 31.2%; Score 5; DB 4; Length 265;  
Best Local Similarity 100.0%; Pred. No. 2.2e+02;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 3 YGRDD 7  
| | | | |  
DB 259 YGRDD 263

RESULT 28  
PCT-US94-10465-3  
; Sequence 3, Application PC/TUS9410465  
; GENERAL INFORMATION:  
; APPLICANT: MEISSNER, ET AL.  
; TITLE OF INVENTION: Human Inositol Monophosphatase H1  
; NUMBER OF SEQUENCES: 3  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: CARILLA, BYRNE, BAIN, GILFILLAN,  
; STREET: 6 BECKER FARM ROAD  
; CITY: ROSELAND  
; STATE: NEW JERSEY  
; COUNTRY: USA  
; ZIP: 07068  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: 3.5 INCH DISKETTE  
; COMPUTER: IBM PS/2

OPERATING SYSTEM: MS-DOS  
; SOFTWARE: WORD PERFECT 5.1  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: PCT/US94/10465  
; FILING DATE: concurrently  
; CLASSIFICATION:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER:  
; FILING DATE:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: FERRARO, GREGORY D.  
; REGISTRATION NUMBER: 36,134  
; REFERENCE/DOCKET NUMBER: 325800-127  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 201-994-1700  
; TELEFAX: 201-994-1744  
; INFORMATION FOR SEQ ID NO: 3:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 265 AMINO ACIDS  
; TYPE: AMINO ACID  
; STRANDEDNESS:  
; TOPOLOGY: LINEAR  
; MOLECULE TYPE: PROTEIN  
; PCT-US94-10465-3  
Query Match 31.2%; Score 5; DB 5; Length 265;  
Best Local Similarity 100.0%; Pred. No. 2.2e+02;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 3 YGRDD 7  
| | | | |  
DB 259 YGRDD 263

RESULT 29  
US-09-134-001C-5291  
; Sequence 5291, Application US/09134001C  
; Patent No. 6380370  
; GENERAL INFORMATION:  
; APPLICANT: Lynn Doucette-Stamm et al  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS  
; FILE REFERENCE: GTC-007  
; CURRENT APPLICATION NUMBER: US/09/134,001C  
; CURRENT FILING DATE: 1998-08-13  
; PRIOR APPLICATION NUMBER: 60/064,964  
; PRIOR FILING DATE: 1997-11-08  
; PRIOR APPLICATION NUMBER: US 60/055,779  
; PRIOR FILING DATE: 1997-08-14  
; NUMBER OF SEQ ID NOS: 5674  
; SEQ ID NO 5291  
; LENGTH: 277  
; TYPE: PRT  
; ORGANISM: Staphylococcus epidermidis  
; US-09-134-001C-5291  
Query Match 31.2%; Score 5; DB 4; Length 277;  
Best Local Similarity 100.0%; Pred. No. 2.2e+02;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 9 DRDPS 13  
| | | | |  
DB 237 DRDPS 241

RESULT 30  
US-09-091-952A-17  
; Sequence 17, Application US/09091952A  
; Patent No. 6458532  
; GENERAL INFORMATION:  
; APPLICANT: Detera-Wedleigh, Sevilla D.  
; Gershon, Elliot S.  
; Badner, Judith A.

```
Goldin, Lynn R.
Berrettini, Wade H.
Yoshikawa, Takeo
Sanders, Alan R.
Baterling, Lisa E.
TITLE OF INVENTION: Chromosomal Markers and Diagnostic
Tests for Manic-Depressive Illness
NUMBER OF SEQUENCES: 197
CORRESPONDENCE ADDRESS:
ADDRESS: Townsend and Townsend and Crew LLP
STREET: Two Embarcadero Center, Eighth Floor
CITY: San Francisco
STATE: CA
COUNTRY: USA
ZIP: 94111-3834
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSeq for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/091,952A
FILING DATE: 19-Apr-1999
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 60/029,278
FILING DATE: 28-OCT-1996
APPLICATION NUMBER: PCT/US97/19381
FILING DATE: 28-OCT-1997
ATTORNEY/AGENT INFORMATION:
NAME: Smith, Timothy L.
REGISTRATION NUMBER: 35,367
REFERENCE/DOCKET NUMBER: 015280-297100US
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 576-0200
TELEFAX: (415) 576-0300
INFORMATION FOR SEQ ID NO: 17:
SEQUENCE CHARACTERISTICS:
LENGTH: 288 amino acids
TYPE: amino acid
STRANDEDNESS: <Unknown>
TOPOLOGY: linear
MOLECULE TYPE: protein
FEATURE:
NAME/KEY: mat.peptide
LOCATION: 1...288
OTHER INFORMATION: IMP.18p myo-inositol monophosphatase
US-09-091-952A-17

Query Match 31.2%; Score 5; DB 4; Length 288;
Best Local Similarity 100.0%; Pred. No. 2.3e+02;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 3 YGRDP 7
Db 282 YGRDP 286

RESULT 31
US-09-103-664A-5
Sequence 5, Application US/09103664A
Patent No. 6458358
GENERAL INFORMATION:
APPLICANT: University of Kentucky Research Foundation
APPLICANT: Timoney, John
APPLICANT: Artushin, Sergey
TITLE OF INVENTION: Compounds Encoding the Protective M-Like Protein of Streptococcus
TITLE OF INVENTION: and Assays Therefor
FILE REFERENCE: 50229-212
CURRENT APPLICATION NUMBER: US/09/103,664A
CURRENT FILING DATE: 1998-06-23
```

```
NUMBER OF SEQ ID NOS: 10
SOFTWARE: PatentIn version 3.0
SEQ ID NO 5
LENGTH: 294
TYPE: PRT
ORGANISM: Streptococcus equi
US-09-103-664A-5

Query Match 31.2%; Score 5; DB 4; Length 294;
Best Local Similarity 100.0%; Pred. No. 2.4e+02;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 3 YGRDP 7
Db 61 YGRDP 65

RESULT 32
US-09-230-637-30
Sequence 30, Application US/09230637
Patent No. 6264958
GENERAL INFORMATION:
APPLICANT: Hayward, Gary
APPLICANT: Nicholas, John
APPLICANT: Hardwick, J. Marie
TITLE OF INVENTION: No. 6264958el Genes of Kaposi's Sarcoma
TITLE OF INVENTION: Associated Herpesvirus
FILE REFERENCE: 1107.78372
CURRENT APPLICATION NUMBER: US/09/230,637
CURRENT FILING DATE: 1999-11-23
PRIOR APPLICATION NUMBER: 60/022,591
PRIOR FILING DATE: 1996-07-25
PRIOR APPLICATION NUMBER: PCT US 97/12931
PRIOR FILING DATE: 1997-07-24
NUMBER OF SEQ ID NOS: 62
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 30
LENGTH: 333
TYPE: PRT
ORGANISM: Kaposi's sarcoma-associated herpes-like virus
US-09-230-637-30

Query Match 31.2%; Score 5; DB 4; Length 333;
Best Local Similarity 100.0%; Pred. No. 2.6e+02;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 4 GRDDN 8
Db 248 GRDDN 252

RESULT 33
US-09-334-601-13
Sequence 13, Application US/09334601
Patent No. 6280989
GENERAL INFORMATION:
APPLICANT: Kapitonov, Dmitri
APPLICANT: Yu, Robert
TITLE OF INVENTION: NOVEL STALYITRANSFERASES
FILE REFERENCE: VCUIP-6
CURRENT APPLICATION NUMBER: US/09/334,601
CURRENT FILING DATE: 1999-06-17
NUMBER OF SEQ ID NOS: 94
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 13
LENGTH: 336
TYPE: PRT
ORGANISM: Homo sapiens
US-09-334-601-13

Query Match 31.2%; Score 5; DB 4; Length 336;
Best Local Similarity 100.0%; Pred. No. 2.6e+02;
```

Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 2 GYGR 6  
      |||||  
Db 130 GYGR 134

RESULT 34  
US-09-739-455-6  
; Sequence 6, Application US/09739455  
; Patent No. 6413756  
; GENERAL INFORMATION:  
; APPLICANT: YAN, Chunhua et al  
; TITLE OF INVENTION: ISOLATED HUMAN KINASE PROTEINS, NUCLEIC  
; TITLE OF INVENTION: ACID MOLECULES ENCODING HUMAN KINASE PROTEINS, AND USES  
; FILE REFERENCE: THEREOF  
; CURRENT APPLICATION NUMBER: US/09/739,455  
; CURRENT FILING DATE: 2000-12-19  
; NUMBER OF SEQ ID NOS: 23  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 6  
; LENGTH: 340  
; TYPE: PRT  
; ORGANISM: Dictyostelium discoideum  
US-09-739-455-6

Query Match 31.2%; Score 5; DB 4; Length 340;  
Best Local Similarity 100.0%; Pred. No. 2.7e+02;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 7 DNDRD 11  
      |||||  
Db 79 DNDRD 83

RESULT 35  
US-09-739-455-16  
; Sequence 16, Application US/09739455  
; Patent No. 6413756  
; GENERAL INFORMATION:  
; APPLICANT: YAN, Chunhua et al  
; TITLE OF INVENTION: ISOLATED HUMAN KINASE PROTEINS, NUCLEIC  
; TITLE OF INVENTION: ACID MOLECULES ENCODING HUMAN KINASE PROTEINS, AND USES  
; FILE REFERENCE: THEREOF  
; CURRENT APPLICATION NUMBER: US/09/739,455  
; CURRENT FILING DATE: 2000-12-19  
; NUMBER OF SEQ ID NOS: 23  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 16  
; LENGTH: 340  
; TYPE: PRT  
; ORGANISM: Dictyostelium discoideum  
US-09-739-455-16

Query Match 31.2%; Score 5; DB 4; Length 340;  
Best Local Similarity 100.0%; Pred. No. 2.7e+02;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 7 DNDRD 11  
      |||||  
Db 79 DNDRD 83

RESULT 36  
US-08-453-924-10  
; Sequence 10, Application US/08453924  
; Patent No. 5608152  
; GENERAL INFORMATION:  
; APPLICANT: Kridl, Jean C.  
; APPLICANT: Knauf, Vic C.  
; TITLE OF INVENTION: Seed-Specific Transcriptional Regulation

NUMBER OF SEQUENCES: 17  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: weil, Gotshal & Manges  
STREET: 2882 Sand Hill Road, Ste. 280  
CITY: Menlo Park  
STATE: CA  
COUNTRY: U.S.A.

ZIP: 94025  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/453,924  
FILING DATE: 30-MAY-1995

CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 07/742,834  
FILING DATE: 08-AUGUST-1991  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 07/550,804  
FILING DATE: 09-JULY-1990  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 07/147,781  
FILING DATE: 25-JANUARY-1988  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 07/078,538  
FILING DATE: 28-JULY-1987  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 06/891,529  
FILING DATE: 31-JULY-1986  
ATTORNEY/AGENT INFORMATION:  
NAME: Rae-Ventley, Barbara  
REGISTRATION NUMBER: 32,750  
REFERENCE/DOCKET NUMBER: CGNE-037/04US  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (415) 926-6200  
TELEFAX: (415) 854-3713  
INFORMATION FOR SEQ ID NO: 10:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 344 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-08-453-924-10

Query Match 31.2%; Score 5; DB 1; Length 344;  
Best Local Similarity 100.0%; Pred. No. 2.7e+02;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VGGR 5  
      |||||  
Db 292 VGGR 296

RESULT 37  
US-09-325-932A-143  
; Sequence 143, Application US/09325932A  
; Patent No. 6451604  
; GENERAL INFORMATION:  
; APPLICANT: Flinn, Barry  
; APPLICANT: Laaham, Annette  
; TITLE OF INVENTION: Compositions affecting programmed cell  
; TITLE OF INVENTION: death and their use in the modification of forestry plant devel  
; FILE REFERENCE: 1022  
; CURRENT APPLICATION NUMBER: US/09/325,932A  
; CURRENT FILING DATE: 1999-06-04  
; NUMBER OF SEQ ID NOS: 206  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 143  
; LENGTH: 386  
; TYPE: PRT



ORGANISM: Eucalyptus grandis  
US-09-325-932A-143

Query Match 31.2%; Score 5; DB 4; Length 386;  
Best Local Similarity 100.0%; Pred. No. 3e+02;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VGYGR 5  
Db 310 VGYGR 314

RESULT 38  
US-07-995-657-6  
Sequence 6, Application US/07995657  
Patent No. 5443974  
GENERAL INFORMATION:  
APPLICANT: Hitz, William D.  
APPLICANT: Yadav, Narendra S.  
TITLE OF INVENTION: Nucleotide Sequence of  
TITLE OF INVENTION: Soybean Stearoyl-ACP  
TITLE OF INVENTION: Desaturase cDNA  
NUMBER OF SEQUENCES: 7  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: E. I. du Pont de Nemours  
ADDRESS: and Company  
STREET: 1007 Market Street  
CITY: Wilmington  
STATE: Delaware  
COUNTRY: U.S.A.  
ZIP: 19898  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette, 3.50 inch, 1.0 MB  
COMPUTER: Macintosh  
OPERATING SYSTEM: Macintosh System, 6.0  
SOFTWARE: Microsoft Word, 4.0  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/07/995.657  
FILING DATE: 19921211  
FILING DATE: C) CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER:  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: GEIGER, KATHLEEN W.  
REGISTRATION NUMBER: 35,880  
REFERENCE/DOCKET NUMBER: BB\_1022-B  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 302-892-8112  
TELEFAX: 302-892-7949  
TELEX: 835420  
INFORMATION FOR SEQ ID NO: 6:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 391 amino acids  
TYPE: AMINO ACID  
STRANDEDNESS: unknown  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-07-995-657-6  
Query Match 31.2%; Score 5; DB 1; Length 391;  
Best Local Similarity 100.0%; Pred. No. 3e+02;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 4 GRDDN 8  
Db 301 GRDDN 305  
RESULT 39  
US-08-474-587-6  
Sequence 6, Application US/08474587

Patent No. 5760206  
GENERAL INFORMATION:  
APPLICANT: Hitz, William D.  
APPLICANT: Yadav, Narendra S.  
APPLICANT: Perez Grau, Luis  
TITLE OF INVENTION: Nucleotide Sequence of  
Patent No. 5760206  
TITLE OF INVENTION: Soybean Stearoyl-ACP  
TITLE OF INVENTION: Desaturase cDNA  
NUMBER OF SEQUENCES: 12  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: E. I. du Pont de Nemours  
ADDRESS: and Company  
STREET: 1007 Market Street  
CITY: Wilmington  
STATE: Delaware  
COUNTRY: U.S.A.  
ZIP: 19898  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette, 3.50 inch, 1.0 MB  
COMPUTER: Macintosh  
OPERATING SYSTEM: Macintosh System, 6.0  
SOFTWARE: Microsoft Word, 4.0  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/474.587  
FILING DATE:  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER:  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: SIEGEL, BARBARA C.  
REGISTRATION NUMBER: 30,684  
REFERENCE/DOCKET NUMBER: BB\_1022-C  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 302-992-4931  
TELEFAX: 302-773-0164  
TELEX: 835420  
INFORMATION FOR SEQ ID NO: 6:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 391 amino acids  
TYPE: amino acid  
STRANDEDNESS: unknown  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-474-587-6  
Query Match 31.2%; Score 5; DB 1; Length 391;  
Best Local Similarity 100.0%; Pred. No. 3e+02;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 4 GRDDN 8  
Db 301 GRDDN 305  
RESULT 40  
US-08-539-798-3  
Sequence 3, Application US/08539798  
Patent No. 5614400  
GENERAL INFORMATION:  
APPLICANT: CAHOON, Edgar B.  
APPLICANT: OHMROGE, John B.  
TITLE OF INVENTION: Methods and Compositions Relating to  
TITLE OF INVENTION: Plant 6-Delta Palmitoyl-Acyl Carrier Protein Desaturase  
NUMBER OF SEQUENCES: 12  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Pioneer Hi-Bred International, Inc.  
STREET: 700 Capital Square, 400 Locust Street  
CITY: Des Moines  
STATE: Iowa  
COUNTRY: US  
ZIP: 50309

```

; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/539,798
; FILING DATE:
; CLASSIFICATION: 800
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/329,560
; FILING DATE: 26-OCT-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Yates, Michael E.
; REGISTRATION NUMBER: 36,063
; REFERENCE/DOCKET NUMBER: 0284US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (515) 248-4800
; TELEFAX: (515) 248-4844
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 396 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
;
US-08-539-798-3

```

```

Query Match      31.2%; Score 5; DB 1; Length 396;
Best Local Similarity 100.0%; Pred.No. 3e+02;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      4 GRDDN 8
      |||||
Db      306 GRDDN 310

```

Search completed: April 9, 2003, 13:16:36  
 Job time : 14.9836 secs

















```
969 4 25.0 197 10 US-09-738-973-206 Sequence 206, App
970 4 25.0 197 10 US-09-803-661-5 Sequence 5, Appl1
971 4 25.0 198 9 US-09-984-245-122 Sequence 122, App
972 4 25.0 198 9 US-09-984-271-183 Sequence 183, App
973 4 25.0 198 9 US-09-984-271-220 Sequence 220, App
974 4 25.0 198 9 US-09-966-262-122 Sequence 122, App
975 4 25.0 198 9 US-09-983-966-122 Sequence 122, App
976 4 25.0 202 9 US-09-738-626-5797 Sequence 5797, App
977 4 25.0 203 10 US-09-781-0808-2 Sequence 2, Appl1
978 4 25.0 204 9 US-09-252-150-15 Sequence 15, Appl1
979 4 25.0 204 9 US-09-738-626-5051 Sequence 5051, App
980 4 25.0 204 9 US-09-738-626-5390 Sequence 5390, App
981 4 25.0 205 10 US-09-925-297-775 Sequence 775, App
982 4 25.0 206 10 US-09-800-528-8 Sequence 8, Appl1
983 4 25.0 209 9 US-10-076-785-16 Sequence 16, Appl1
984 4 25.0 209 10 US-09-393-634-13 Sequence 13, Appl1
985 4 25.0 210 10 US-09-189-833B-10 Sequence 10, Appl1
986 4 25.0 210 10 US-09-902-705-10 Sequence 10, Appl1
987 4 25.0 211 9 US-10-046-938-2 Sequence 2, Appl1
988 4 25.0 211 9 US-09-738-626-4101 Sequence 4101, App
989 4 25.0 212 9 US-10-076-785-20 Sequence 20, Appl1
990 4 25.0 212 9 US-09-738-626-6234 Sequence 6234, App
991 4 25.0 212 9 US-10-146-474-4 Sequence 4, Appl1
992 4 25.0 212 10 US-09-815-242-12923 Sequence 12923, A
993 4 25.0 212 10 US-09-815-242-12949 Sequence 12949, A
994 4 25.0 212 10 US-09-815-242-13097 Sequence 13097, A
995 4 25.0 212 10 US-09-815-242-13122 Sequence 13122, A
996 4 25.0 212 10 US-09-764-864-937 Sequence 937, App
997 4 25.0 213 10 US-09-924-256A-64 Sequence 64, Appl1
998 4 25.0 214 10 US-09-924-256A-44 Sequence 44, Appl1
999 4 25.0 214 10 US-09-924-256A-74 Sequence 74, Appl1
1000 4 25.0 214 10 US-09-924-256A-78 Sequence 78, Appl1
```

## ALIGNMENTS

```
RESULT 1
US-10-020-441-3
; Sequence 3, Application US/10020441
; Publication No. US20020182224A1
; GENERAL INFORMATION:
; APPLICANT: Doenhoff, Michael
; APPLICANT: Sayers, Jon
; TITLE OF INVENTION: SCHISTOSOMIASIS VACCINE
; FILE REFERENCE: Doenhoff
; CURRENT APPLICATION NUMBER: US/10/020,441
; PRIOR FILING DATE: 2001-12-18
; PRIOR APPLICATION NUMBER: US/09/413,810
; NUMBER OF SEQ ID NOS: 3
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 3
; LENGTH: 16
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: protein
US-10-020-441-3
```

```
Query Match 100.0%; Score 16; DB 9; Length 16;
Best Local Similarity 100.0%; Pred.No. 5.3e-11;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
Qy 1 VGYGRDNDNDPDSRKN 16
Db 1 VGYGRDNDNDPDSRKN 16
```

```
RESULT 2
US-10-020-441-2
; Sequence 2, Application US/10020441
; Publication No. US20020182224A1
```

```
; GENERAL INFORMATION:
; APPLICANT: Doenhoff, Michael
; APPLICANT: Sayers, Jon
; TITLE OF INVENTION: SCHISTOSOMIASIS VACCINE
; FILE REFERENCE: Doenhoff
; CURRENT APPLICATION NUMBER: US/10/020,441
; PRIOR FILING DATE: 2001-12-18
; PRIOR APPLICATION NUMBER: US/09/413,810
; NUMBER OF SEQ ID NOS: 3
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2
; LENGTH: 106
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: protein
US-10-020-441-2
```

```
Query Match 100.0%; Score 16; DB 9; Length 106;
Best Local Similarity 100.0%; Pred.No. 2.7e-10;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
Qy 1 VGYGRDNDNDPDSRKN 16
Db 85 VGYGRDNDNDPDSRKN 100
```

```
RESULT 3
US-09-864-761-47769
; Sequence 47769, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharon G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
; FILE REFERENCE: Aecomica-X-1
; CURRENT APPLICATION NUMBER: US/09/864,761
; PRIOR FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/632,366
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: GB 24263.6
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US01/00666
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00664
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00669
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00665
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00668
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00663
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00662
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00661
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00670
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: US 60/234,687
```

```

; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: US 09/608,408
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/774,203
; PRIOR FILING DATE: 2001-01-29
; NUMBER OF SEQ ID NOS: 49117
; SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 47769
; LENGTH: 30
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AC011619.1
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 13
; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 0.98
US-09-864-761-47769

Query Match      37.5%; Score 6; DB 10; Length 30;
Best Local Similarity 100.0%; Pred. No. 3.1;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      8 NDRDPS 13
Db      6 NDRDPS 11

RESULT 4
US-09-852-797-113
; Sequence 113, Application US/09852797
; Patent No. US20020172994A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: 28 Human Secreted Proteins
; FILE REFERENCE: P2003P2
; CURRENT APPLICATION NUMBER: US/09/852,797
; PRIOR FILING DATE: 2001-05-11
; PRIOR APPLICATION NUMBER: 60/265,583
; PRIOR FILING DATE: 2001-02-02
; PRIOR APPLICATION NUMBER: 09/152,060
; PRIOR FILING DATE: 1998-09-11
; PRIOR APPLICATION NUMBER: PCT/US98/04858
; PRIOR FILING DATE: 1998-03-12
; PRIOR APPLICATION NUMBER: 60/040,762
; PRIOR FILING DATE: 1997-03-14
; PRIOR APPLICATION NUMBER: 60/040,710
; PRIOR FILING DATE: 1997-03-14
; PRIOR APPLICATION NUMBER: 60/050,934
; PRIOR FILING DATE: 1997-05-30
; PRIOR APPLICATION NUMBER: 60/048,100
; PRIOR FILING DATE: 1997-05-30
; PRIOR APPLICATION NUMBER: 60/048,357
; PRIOR FILING DATE: 1997-05-30
; PRIOR APPLICATION NUMBER: 60/048,189
; PRIOR FILING DATE: 1997-05-30
; PRIOR APPLICATION NUMBER: 60/057,765
; PRIOR FILING DATE: 1997-09-05
; PRIOR APPLICATION NUMBER: 60/048,970
; PRIOR FILING DATE: 1997-06-06
; PRIOR APPLICATION NUMBER: 60/068,368
; PRIOR FILING DATE: 1997-12-19
; NUMBER OF SEQ ID NOS: 118
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 113
; LENGTH: 18
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-852-797-113

Query Match      31.2%; Score 5; DB 9; Length 18;
Best Local Similarity 100.0%; Pred. No. 23;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      2 GYGRD 6
Db      4 GYGRD 6
```

```

; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: US 09/608,408
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/774,203
; PRIOR FILING DATE: 2001-01-29
; NUMBER OF SEQ ID NOS: 49117
; SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 47769
; LENGTH: 30
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AC011619.1
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 13
; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 0.98
US-09-864-761-47769

Query Match      37.5%; Score 6; DB 10; Length 30;
Best Local Similarity 100.0%; Pred. No. 3.1;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      8 NDRDPS 13
Db      6 NDRDPS 11

RESULT 4
US-09-852-797-113
; Sequence 113, Application US/09852797
; Patent No. US20020172994A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: 28 Human Secreted Proteins
; FILE REFERENCE: P2003P2
; CURRENT APPLICATION NUMBER: US/09/852,797
; PRIOR FILING DATE: 2001-05-11
; PRIOR APPLICATION NUMBER: 60/265,583
; PRIOR FILING DATE: 2001-02-02
; PRIOR APPLICATION NUMBER: 09/152,060
; PRIOR FILING DATE: 1998-09-11
; PRIOR APPLICATION NUMBER: PCT/US98/04858
; PRIOR FILING DATE: 1998-03-12
; PRIOR APPLICATION NUMBER: 60/040,762
; PRIOR FILING DATE: 1997-03-14
; PRIOR APPLICATION NUMBER: 60/040,710
; PRIOR FILING DATE: 1997-03-14
; PRIOR APPLICATION NUMBER: 60/050,934
; PRIOR FILING DATE: 1997-05-30
; PRIOR APPLICATION NUMBER: 60/048,100
; PRIOR FILING DATE: 1997-05-30
; PRIOR APPLICATION NUMBER: 60/048,357
; PRIOR FILING DATE: 1997-05-30
; PRIOR APPLICATION NUMBER: 60/048,189
; PRIOR FILING DATE: 1997-05-30
; PRIOR APPLICATION NUMBER: 60/057,765
; PRIOR FILING DATE: 1997-09-05
; PRIOR APPLICATION NUMBER: 60/048,970
; PRIOR FILING DATE: 1997-06-06
; PRIOR APPLICATION NUMBER: 60/068,368
; PRIOR FILING DATE: 1997-12-19
; NUMBER OF SEQ ID NOS: 118
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 113
; LENGTH: 18
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-853-161-113

Query Match      31.2%; Score 5; DB 10; Length 18;
Best Local Similarity 100.0%; Pred. No. 23;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      2 GYGRD 6
Db      4 GYGRD 8

RESULT 6
US-09-852-659A-113
; Sequence 113, Application US/09852659A
; Patent No. US20020077287A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: 28 Human Secreted Proteins
; FILE REFERENCE: P2003P4
; CURRENT APPLICATION NUMBER: US/09/852,659A
; PRIOR FILING DATE: 2001-05-11
; PRIOR APPLICATION NUMBER: 60/265,583
; PRIOR FILING DATE: 2001-02-02
; PRIOR APPLICATION NUMBER: 09/152,060
; PRIOR FILING DATE: 1998-09-11
; PRIOR APPLICATION NUMBER: PCT/US98/04858
; PRIOR FILING DATE: 1998-03-12
; PRIOR APPLICATION NUMBER: 60/040,762
; PRIOR FILING DATE: 1997-03-14
```

```
; PRIOR APPLICATION NUMBER: 60/040,710
; PRIOR FILING DATE: 1997-03-14
; PRIOR APPLICATION NUMBER: 60/050,934
; PRIOR FILING DATE: 1997-05-30
; PRIOR APPLICATION NUMBER: 60/048,100
; PRIOR FILING DATE: 1997-05-30
; PRIOR APPLICATION NUMBER: 60/048,357
; PRIOR FILING DATE: 1997-05-30
; PRIOR APPLICATION NUMBER: 60/048,189
; PRIOR FILING DATE: 1997-05-30
; PRIOR APPLICATION NUMBER: 60/057,765
; PRIOR FILING DATE: 1997-09-05
; PRIOR APPLICATION NUMBER: 60/048,970
; PRIOR FILING DATE: 1997-06-06
; PRIOR APPLICATION NUMBER: 60/068,368
; PRIOR FILING DATE: 1997-12-19
; NUMBER OF SEQ ID NOS: 121
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 113
; LENGTH: 18
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-852-659A-113

Query Match      31.2%; Score 5; DB 10; Length 18;
Best Local Similarity 100.0%; Pred. No. 23;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      2 GYGRD 6
      |||||
Db      4 GYGRD 8

RESULT 7
US-09-983-802-493
; Sequence 493, Application US/09983802
; Publication No. US20030022185A1
; GENERAL INFORMATION:
; APPLICANT: Fischer et al.
; TITLE OF INVENTION: 123 Human Secreted Proteins
; FILE REFERENCE: P2010P1
; CURRENT APPLICATION NUMBER: US/09/983,802
; PRIOR FILING DATE: 2001-10-25
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 09/227,357
; PRIOR FILING DATE: EARLIER FILING DATE: 1999-01-08
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: PCT/US98/13684
; PRIOR FILING DATE: EARLIER FILING DATE: 1998-07-07
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/051,926
; PRIOR FILING DATE: EARLIER FILING DATE: 1997-07-08
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/052,793
; PRIOR FILING DATE: EARLIER FILING DATE: 1997-07-08
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/051,925
; PRIOR FILING DATE: EARLIER FILING DATE: 1997-07-08
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/051,929
; PRIOR FILING DATE: EARLIER FILING DATE: 1997-07-08
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/052,803
; PRIOR FILING DATE: EARLIER FILING DATE: 1997-07-08
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/052,732
; PRIOR FILING DATE: EARLIER FILING DATE: 1997-07-08
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/051,931
; PRIOR FILING DATE: EARLIER FILING DATE: 1997-07-08
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/051,932
; PRIOR FILING DATE: EARLIER FILING DATE: 1997-07-08
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/051,916
; PRIOR FILING DATE: EARLIER FILING DATE: 1997-07-08
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/051,930
; PRIOR FILING DATE: EARLIER FILING DATE: 1997-07-08
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/051,918
; PRIOR FILING DATE: EARLIER FILING DATE: 1997-07-08
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/051,920
; PRIOR FILING DATE: EARLIER FILING DATE: 1997-07-08
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/052,733
; PRIOR FILING DATE: EARLIER FILING DATE: 1997-07-08
```

```
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/052,795
; PRIOR FILING DATE: EARLIER FILING DATE: 1997-07-08
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/051,919
; PRIOR FILING DATE: EARLIER FILING DATE: 1997-07-08
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/051,928
; PRIOR FILING DATE: EARLIER FILING DATE: 1997-07-08
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/055,722
; PRIOR FILING DATE: EARLIER FILING DATE: 1997-08-18
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/055,723
; PRIOR FILING DATE: EARLIER FILING DATE: 1997-08-18
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/055,948
; PRIOR FILING DATE: EARLIER FILING DATE: 1997-08-18
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/055,949
; PRIOR FILING DATE: EARLIER FILING DATE: 1997-08-18
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/055,953
; PRIOR FILING DATE: EARLIER FILING DATE: 1997-08-18
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/055,950
; PRIOR FILING DATE: EARLIER FILING DATE: 1997-08-18
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/055,947
; PRIOR FILING DATE: EARLIER FILING DATE: 1997-08-18
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/055,964
; PRIOR FILING DATE: EARLIER FILING DATE: 1997-08-18
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/056,360
; PRIOR FILING DATE: EARLIER FILING DATE: 1997-08-18
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/055,954
; PRIOR FILING DATE: EARLIER FILING DATE: 1997-08-18
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/058,785
; PRIOR FILING DATE: EARLIER FILING DATE: 1997-09-12
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/058,664
; PRIOR FILING DATE: EARLIER FILING DATE: 1997-09-12
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/058,660
; PRIOR FILING DATE: EARLIER FILING DATE: 1997-09-12
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/058,661
; PRIOR FILING DATE: EARLIER FILING DATE: 1997-09-12
; NUMBER OF SEQ ID NOS: 672
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 493
; LENGTH: 24
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-983-802-493

Query Match      31.2%; Score 5; DB 9; Length 24;
Best Local Similarity 100.0%; Pred. No. 29;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      10 RDPSP 14
      |||||
Db      18 RDPSP 22

RESULT 8
US-09-864-761-42580
; Sequence 42580, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharon G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wenheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
; FILE REFERENCE: Aesomica-X-1
; CURRENT APPLICATION NUMBER: US/09/864,761
; PRIOR FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-05-26
```

```
; PRIOR APPLICATION NUMBER: US 09/632,366
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: GB 24263.6
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US01/00666
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00664
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00669
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00665
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00668
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00663
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00662
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00661
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00670
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: US 09/608,408
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/774,203
; PRIOR FILING DATE: 2001-01-29
; NUMBER OF SEQ ID NOS: 49117
; SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 42580
; LENGTH: 51
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AC008583.4
; OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 3.3
; OTHER INFORMATION: SWISSPROT HIT: P14547, EVALU6 6.10e+00
US-09-864-761-42580

Query Match          31.2%; Score 5; DB 10; Length 51;
Best Local Similarity 100.0%; Pred. No. 56;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
Qy      12 PSRKN 16
      |||||
Db      21 PSRKN 25

RESULT 9
US-09-864-761-40351
; Sequence 40351, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharon G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
; FILE REFERENCE: Aecmica-X-1
; CURRENT APPLICATION NUMBER: US/09/864,761
; PRIOR FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/632,366
; PRIOR FILING DATE: 2000-08-03
```

```
; PRIOR APPLICATION NUMBER: GB 24263.6
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US01/00666
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00664
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00669
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00665
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00668
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00663
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00662
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00661
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00670
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: US 09/608,408
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/774,203
; PRIOR FILING DATE: 2001-01-29
; NUMBER OF SEQ ID NOS: 49117
; SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 40351
; LENGTH: 52
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AL136984.7
; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 2.6
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 2.2
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 2.2
; OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 2.1
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 2.2
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 2
; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 2.1
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 2.2
US-09-864-761-40351

Query Match          31.2%; Score 5; DB 10; Length 52;
Best Local Similarity 100.0%; Pred. No. 57;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
Qy      12 PSRKN 16
      |||||
Db      38 PSRKN 42

RESULT 10
US-09-864-761-43987
; Sequence 43987, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharon G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
; FILE REFERENCE: Aecmica-X-1
; CURRENT APPLICATION NUMBER: US/09/864,761
; PRIOR FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 2000-02-04
```

```
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/632,366
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: GB 24263.6
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US01/00666
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00664
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00669
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00665
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00668
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00663
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00662
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00661
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00670
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: US 09/608,408
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/774,203
; PRIOR FILING DATE: 2001-01-29
; NUMBER OF SEQ ID NOS: 49117
; SOFTWARE: Annonax Sequence Listing Engine vers. 1.1
; SEQ ID NO 43987
; LENGTH: 68
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AC002486.1
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 1.1
; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 1
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 1.2
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 0.95
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 0.96
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 1.1
; OTHER INFORMATION: SWISSPROT HIT: P21447, EVALUE 5.00e-15
US-09-864-761-43987

Query Match          31.2%; Score 5; DB 10; Length 68;
Best Local Similarity 100.0%; Pred. No. 72;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      3 YGRDD 7
      |||||
      40 YGRDD 44

RESULT 11
US-10-050-786-9
; Sequence 9, Application US/10050786
; Patent No. US2002015539A1
; GENERAL INFORMATION:
; APPLICANT: Ruben et al.
; TITLE OF INVENTION: Calcium Channel Polynucleotides, Polypeptides, and Antibodies
; FILE REFERENCE: PT013PIC1
; CURRENT APPLICATION NUMBER: US/10/050,786
; PRIOR FILING DATE: 2002-01-18
; CURRENT APPLICATION NUMBER: US 09/774,028
; PRIOR FILING DATE: 2001-01-31
; PRIOR APPLICATION NUMBER: PCT/US00/20392
```

```
; PRIOR FILING DATE: 2000-07-27
; PRIOR APPLICATION NUMBER: US 60/145,958
; PRIOR FILING DATE: 1999-07-28
; PRIOR APPLICATION NUMBER: US 60/149,446
; PRIOR FILING DATE: 1999-08-18
; PRIOR APPLICATION NUMBER: US 60/189,064
; PRIOR FILING DATE: 2000-03-14
; NUMBER OF SEQ ID NOS: 12
; SOFTWARE: PatentIn Ver. 3.1
; SEQ ID NO 9
; LENGTH: 103
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-050-786-9

Query Match          31.2%; Score 5; DB 9; Length 103;
Best Local Similarity 100.0%; Pred. No. 1e+02;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      9 DRDPs 13
      |||||
      70 DRDPs 74

RESULT 12
US-09-866-050A-179
; Sequence 179, Application US/09866050A
; Publication No. US2003004047A1
; GENERAL INFORMATION:
; APPLICANT: Watson, James D.
; APPLICANT: Strachan, Lorna
; APPLICANT: Sleeman, Matthew
; APPLICANT: Onrust, Rene
; APPLICANT: Murison, James G.
; APPLICANT: Kumble, Krishanand D.
; TITLE OF INVENTION: Compositions Isolated From Skin Cells
; FILE REFERENCE: 11000.1011c4U
; CURRENT APPLICATION NUMBER: US/09/866,050A
; CURRENT FILING DATE: 2001-05-24
; NUMBER OF SEQ ID NOS: 725
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 179
; LENGTH: 123
; TYPE: PRT
; ORGANISM: Mouse
US-09-866-050A-179

Query Match          31.2%; Score 5; DB 9; Length 123;
Best Local Similarity 100.0%; Pred. No. 1.2e+02;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      12 PSRKN 16
      |||||
      113 PSRKN 117

RESULT 13
US-09-867-550-1718
; Sequence 1718, Application US/09867550
; Patent No. US20020082206A1
; GENERAL INFORMATION:
; APPLICANT: Leach, Martin D.
; APPLICANT: Mehrahan, Fuad,
; APPLICANT: Conley, Pamela
; APPLICANT: Law, Debbie
; APPLICANT: Topper, James
; TITLE OF INVENTION: No. US20020082206A1 Polynucleotides from Atherogenic Cells and
; FILE REFERENCE: 21402-013 (Cura-313)
; CURRENT APPLICATION NUMBER: US/09/867,550
; CURRENT FILING DATE: 2001-09-20
; PRIOR APPLICATION NUMBER: USSN 60/208,427
```

```
; PRIOR FILING DATE: 2000-05-30
; NUMBER OF SEQ ID NOS: 2125
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1718
; LENGTH: 124
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: VARIANT
; LOCATION: (108)
; OTHER INFORMATION: wherein Xaa may be any one of Ala or Pro or Ser or Thr
US-09-867-550-1718

Query Match          31.2%; Score 5; DB 10; Length 124;
Best Local Similarity 100.0%; Pred. No. 1.2e+02;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 12 PSRKN 16
   |||||
Db 83 PSRKN 87

RESULT 14
US-09-873-409-3
; Sequence 3, Application US/09873409
; Patent No. US20020037522A1
; GENERAL INFORMATION:
; APPLICANT: Frank, Markus
; TITLE OF INVENTION: A Gene Encoding a Multidrug Resistance Human P-Glycoprotein
; FILE REFERENCE: 81994/268611
; CURRENT APPLICATION NUMBER: US/09/873,409
; PRIOR FILING DATE: 2001-06-05
; NUMBER OF SEQ ID NOS: 19
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 3
; LENGTH: 131
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-873-409-3

Query Match          31.2%; Score 5; DB 10; Length 131;
Best Local Similarity 100.0%; Pred. No. 1.3e+02;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 YGRDD 7
   |||||
Db 39 YGRDD 43

RESULT 15
US-09-925-299-798
; Sequence 798, Application US/09925299
; Publication No. US20030040617A9
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
; FILE REFERENCE: PA102
; CURRENT APPLICATION NUMBER: US/09/925,299
; PRIOR FILING DATE: 2001-08-10
; PRIOR APPLICATION NUMBER: PCT/US00/05883
; PRIOR FILING DATE: 2000-03-08
; PRIOR APPLICATION NUMBER: 60/124,270
; PRIOR FILING DATE: 1999-03-12
; NUMBER OF SEQ ID NOS: 1556
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 798
; LENGTH: 137
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
```

```
; LOCATION: (8)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-09-925-299-798

Query Match          31.2%; Score 5; DB 9; Length 137;
Best Local Similarity 100.0%; Pred. No. 1.3e+02;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 9 DRDPS 13
   |||||
Db 39 DRDPS 43

RESULT 16
US-09-925-299-798
; Sequence 798, Application US/09925299
; Patent No. US20020055627A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
; FILE REFERENCE: PA102
; CURRENT APPLICATION NUMBER: US/09/925,299
; PRIOR FILING DATE: 2001-08-10
; PRIOR APPLICATION NUMBER: PCT/US00/05883
; PRIOR FILING DATE: 2000-03-08
; PRIOR APPLICATION NUMBER: 60/124,270
; PRIOR FILING DATE: 1999-03-12
; NUMBER OF SEQ ID NOS: 1556
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 798
; LENGTH: 137
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (8)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-09-925-299-798

Query Match          31.2%; Score 5; DB 10; Length 137;
Best Local Similarity 100.0%; Pred. No. 1.3e+02;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 9 DRDPS 13
   |||||
Db 39 DRDPS 43

RESULT 17
US-10-079-623-367
; Sequence 367, Application US/10079623
; Patent No. US20020169302A1
; GENERAL INFORMATION:
; APPLICANT: Havukkala, Ilkka J.
; APPLICANT: Glenn, Matthew
; APPLICANT: Grigor, Murray R.
; APPLICANT: Molenaar, Adrian J.
; TITLE OF INVENTION: Compositions isolated from bovine
; FILE REFERENCE: 11000.10443
; CURRENT APPLICATION NUMBER: US/10/079,623
; CURRENT FILING DATE: 2002-02-19
; NUMBER OF SEQ ID NOS: 370
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 367
; LENGTH: 142
; TYPE: PRT
; ORGANISM: Bovine
US-10-079-623-367

Query Match          31.2%; Score 5; DB 9; Length 142;
Best Local Similarity 100.0%; Pred. No. 1.4e+02;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

Oy 3 YGRDD 7  
|||||  
Db 65 YGRDD 69

## RESULT 18

US-09-761-782-2  
; Sequence 2, Application US/09761782  
; Patent No. US20020037562A1  
; GENERAL INFORMATION:  
; APPLICANT: LIVSHITS, VITALITY  
; APPLICANT: DOROSHENKO, VERA  
; APPLICANT: GORSHKOVA, NATALIYA  
; APPLICANT: BELARYEVA, ALLA  
; APPLICANT: IVANOVSKAYA, LIRINA  
; APPLICANT: KHOURGES, EVGENI  
; APPLICANT: AKHVERDIAN, VALERIY  
; APPLICANT: GUSVATINER, MIKHAIL  
; APPLICANT: KOZLOV, YURY  
; TITLE OF INVENTION: MUTANT LIVH GENE AND METHOD FOR PRODUCING L-VALINE  
; FILE REFERENCE: 202048USO  
; CURRENT APPLICATION NUMBER: US/09/761,782  
; PRIOR FILING DATE: 2001-01-18  
; PRIOR APPLICATION NUMBER: RU 2000101678  
; PRIOR FILING DATE: 2000-01-26  
; NUMBER OF SEQ ID NOS: 8  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 2  
; LENGTH: 163  
; TYPE: PRT  
; ORGANISM: Escherichia coli  
US-09-761-782-2

Query Match 31.2%; Score 5; DB 10; Length 163;  
Best Local Similarity 100.0%; Pred. No. 1.5e+02;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 2 GYGRD 6  
|||||  
Db 95 GYGRD 99

## RESULT 19

US-09-815-242-10023  
; Sequence 10023, Application US/09815242  
; Patent No. US20020061569A1  
; GENERAL INFORMATION:  
; APPLICANT: Haselbeck, Robert  
; APPLICANT: Ohlsen, Karl L.  
; APPLICANT: Zyskind, Judith W.  
; APPLICANT: Wall, Daniel  
; APPLICANT: Trawick, John D.  
; APPLICANT: Carr, Grant J.  
; APPLICANT: Yamamoto, Robert T.  
; APPLICANT: Xu, H. Howard  
; TITLE OF INVENTION: Identification of Essential Genes in  
; TITLE OF INVENTION: Prokaryotes  
; FILE REFERENCE: ELITRA.011A  
; CURRENT APPLICATION NUMBER: US/09/815,242  
; CURRENT FILING DATE: 2001-03-21  
; PRIOR APPLICATION NUMBER: 60/191,078  
; PRIOR FILING DATE: 2000-03-21  
; PRIOR APPLICATION NUMBER: 60/206,848  
; PRIOR FILING DATE: 2000-05-23  
; PRIOR APPLICATION NUMBER: 60/207,727  
; PRIOR FILING DATE: 2000-05-26  
; PRIOR APPLICATION NUMBER: 60/242,578  
; PRIOR FILING DATE: 2000-10-23  
; PRIOR APPLICATION NUMBER: 60/253,625  
; PRIOR FILING DATE: 2000-11-27  
; PRIOR APPLICATION NUMBER: 60/257,931  
; PRIOR FILING DATE: 2000-12-22

; PRIOR APPLICATION NUMBER: 60/269,308  
; PRIOR FILING DATE: 2001-02-16  
; NUMBER OF SEQ ID NOS: 14110  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 10023  
; LENGTH: 163  
; TYPE: PRT  
; ORGANISM: Escherichia coli  
US-09-815-242-10023

Query Match 31.2%; Score 5; DB 10; Length 163;  
Best Local Similarity 100.0%; Pred. No. 1.5e+02;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 2 GYGRD 6  
|||||  
Db 95 GYGRD 99

## RESULT 20

US-09-764-868-841  
; Sequence 841, Application US/09764868  
; Patent No. US20020168711A1  
; GENERAL INFORMATION:  
; APPLICANT: Rosen et al.  
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies  
; FILE REFERENCE: PT232  
; CURRENT APPLICATION NUMBER: US/09/764,868  
; PRIOR FILING DATE: 2001-01-17  
; Prior application data removed - refer to PALM or file wrapper  
; NUMBER OF SEQ ID NOS: 1510  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 841  
; LENGTH: 166  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: SITE  
; LOCATION: (142)  
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids  
; NAME/KEY: SITE  
; LOCATION: (150)  
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids  
; NAME/KEY: SITE  
; LOCATION: (157)  
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids  
US-09-764-868-841

Query Match 31.2%; Score 5; DB 9; Length 166;  
Best Local Similarity 100.0%; Pred. No. 1.6e+02;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 10 RDPGR 14  
|||||  
Db 62 RDPGR 66

## RESULT 21

US-10-214-811-14  
; Sequence 14, Application US/10214811  
; Publication No. US20030023062A1  
; GENERAL INFORMATION:  
; APPLICANT: Donoho, Gregory  
; APPLICANT: Scoville, John  
; APPLICANT: Turner, C. Alexander Jr.  
; APPLICANT: Friedlich, Glenn  
; APPLICANT: Abuin, Alejandro  
; APPLICANT: Zambrowicz, Brian  
; APPLICANT: Sands, Arthur T.  
; TITLE OF INVENTION: No. US20030023062A1 Human Processes and  
; TITLE OF INVENTION: Polynucleotides Encoding the Same  
; FILE REFERENCE: LEX-0132-USA  
; CURRENT APPLICATION NUMBER: US/10/214,811



```
; CURRENT FILING DATE: 2002-08-07
; PRIOR APPLICATION NUMBER: US/09/780,016
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: US 60/181,294
; NUMBER OF SEQ ID NOS: 27
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 14
; LENGTH: 186
; TYPE: PRT
; ORGANISM: homo sapiens
US-10-214-811-14

Query Match          31.2%; Score 5; DB 9; Length 186;
Best Local Similarity 100.0%; Pred. No. 1,7e+02;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      10 RDPSP 14
       |||||
Db       77 RDPSP 81

RESULT 22
US-09-780-016-14
; Sequence 14, Application US/09780016
; Patent No. US20020004591A1
; GENERAL INFORMATION:
; APPLICANT: Donoho, Gregory
; APPLICANT: Scoville, John
; APPLICANT: Turner, C. Alexander Jr.
; APPLICANT: Friedrich, Glenn
; APPLICANT: Abulin, Alejandro
; APPLICANT: Zambrowicz, Brian
; APPLICANT: Sands, Arthur T.
; TITLE OF INVENTION: No. US20020004591A1el Human Proteases and
; FILE OF INVENTION: Polynucleotides Encoding the Same
; CURRENT APPLICATION NUMBER: US/09/780,016
; CURRENT FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: US 60/181,294
; PRIOR FILING DATE: 2000-02-11
; NUMBER OF SEQ ID NOS: 27
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 14
; LENGTH: 186
; TYPE: PRT
; ORGANISM: homo sapiens
US-09-780-016-14

Query Match          31.2%; Score 5; DB 10; Length 186;
Best Local Similarity 100.0%; Pred. No. 1,7e+02;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      10 RDPSP 14
       |||||
Db       77 RDPSP 81

RESULT 23
US-09-908-711-121
; Sequence 121, Application US/09908711
; Patent No. US20020045230A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PA128
; CURRENT APPLICATION NUMBER: US/09/908,711
; CURRENT FILING DATE: 2001-07-20
; PRIOR APPLICATION NUMBER: US01/01360
; PRIOR FILING DATE: 2001-01-17
; PRIOR APPLICATION NUMBER: 09/764,867
; PRIOR FILING DATE: 2001-01-17
; PRIOR APPLICATION NUMBER: US01/01344
```

```
; PRIOR FILING DATE: 2001-01-17
; PRIOR APPLICATION NUMBER: 09/764,892
; PRIOR FILING DATE: 2001-01-17
; PRIOR APPLICATION NUMBER: US01/01345
; PRIOR FILING DATE: 2001-01-17
; PRIOR APPLICATION NUMBER: 09/764,888
; PRIOR FILING DATE: 2001-01-17
; PRIOR APPLICATION NUMBER: US01/01329
; PRIOR FILING DATE: 2001-01-17
; PRIOR APPLICATION NUMBER: 09/764,905
; PRIOR FILING DATE: 2001-01-17
; PRIOR APPLICATION NUMBER: US01/01354
; PRIOR FILING DATE: 2001-01-17
; PRIOR APPLICATION NUMBER: 09/764,891
; PRIOR FILING DATE: 2001-01-17
; PRIOR APPLICATION NUMBER: US01/01339
; PRIOR FILING DATE: 2001-01-17
; PRIOR APPLICATION NUMBER: 09/764,869
; PRIOR FILING DATE: 2001-01-17
; PRIOR APPLICATION NUMBER: US01/01340
; PRIOR FILING DATE: 2001-01-17
; PRIOR APPLICATION NUMBER: 09/764,874
; PRIOR FILING DATE: 2001-01-17
; PRIOR APPLICATION NUMBER: US01/01334
; PRIOR FILING DATE: 2001-01-17
; PRIOR APPLICATION NUMBER: 09/764,898
; PRIOR FILING DATE: 2001-01-17
; PRIOR APPLICATION NUMBER: US01/01320
; PRIOR FILING DATE: 2001-01-17
; PRIOR APPLICATION NUMBER: 09/764,853
; PRIOR FILING DATE: 2001-01-17
; PRIOR APPLICATION NUMBER: US01/01349
; PRIOR FILING DATE: 2001-01-17
; PRIOR APPLICATION NUMBER: 09/764,902
; PRIOR FILING DATE: 2001-01-17
; PRIOR APPLICATION NUMBER: US01/01239
; PRIOR FILING DATE: 2001-01-17
; PRIOR APPLICATION NUMBER: 09/764,870
; PRIOR FILING DATE: 2001-01-17
; PRIOR APPLICATION NUMBER: US01/01348
; PRIOR FILING DATE: 2001-01-17
; PRIOR APPLICATION NUMBER: 09/764,882
; PRIOR FILING DATE: 2001-01-17
; PRIOR APPLICATION NUMBER: US01/01347
; PRIOR FILING DATE: 2001-01-17
; PRIOR APPLICATION NUMBER: 09/764,896
; PRIOR FILING DATE: 2001-01-17
; PRIOR APPLICATION NUMBER: US01/01307
; PRIOR FILING DATE: 2001-01-17
; PRIOR APPLICATION NUMBER: 09/764,864
; PRIOR FILING DATE: 2001-01-17
; PRIOR APPLICATION NUMBER: US01/01341
; PRIOR FILING DATE: 2001-01-17
; PRIOR APPLICATION NUMBER: 09/764,856
; PRIOR FILING DATE: 2001-01-17
; PRIOR APPLICATION NUMBER: US01/01336
; PRIOR FILING DATE: 2001-01-17
; PRIOR APPLICATION NUMBER: 09/764,868
; PRIOR FILING DATE: 2001-01-17
; PRIOR APPLICATION NUMBER: US01/01312
; PRIOR FILING DATE: 2001-01-17
; PRIOR APPLICATION NUMBER: 60/179,065
; PRIOR FILING DATE: 2000-01-31
; PRIOR APPLICATION NUMBER: 60/180,628
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: 60/209,467
; PRIOR FILING DATE: 2000-06-07
; NUMBER OF SEQ ID NOS: 167
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 121
; LENGTH: 190
; TYPE: PRT
; ORGANISM: Homo sapiens
```

US-09-908-711-121

Query Match 31.2% Score 5; DB 10; Length 190;  
Best Local Similarity 100.0%; Pred. No. 1.7e+02;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 11 DPSRK 15

DB 18 DPSRK 22

RESULT 24

US-10-214-811-10  
; Sequence 10, Application US/10214811  
; Publication No. US20030023062A1  
; GENERAL INFORMATION:  
; APPLICANT: Donoho, Gregory  
; APPLICANT: Scoville, John  
; APPLICANT: Turner, C. Alexander Jr.  
; APPLICANT: Friedlich, Glenn  
; APPLICANT: Abuin, Alejandro  
; APPLICANT: Zambrowicz, Brian  
; APPLICANT: Sands, Arthur T.  
; TITLE OF INVENTION: No. US20030023062A1el Human Proteases and  
; TITLE OF INVENTION: Polynucleotides Encoding the Same  
; FILE REFERENCE: LEX-0132-USA  
; CURRENT FILING DATE: US/10/214, 811  
; PRIOR FILING DATE: 2002-08-07  
; PRIOR APPLICATION NUMBER: US/09/780, 016  
; PRIOR FILING DATE: 2001-02-09  
; PRIOR APPLICATION NUMBER: US 60/181,294  
; PRIOR FILING DATE: 2000-02-11  
; NUMBER OF SEQ ID NOS: 27  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 10  
; LENGTH: 211  
; TYPE: PRT  
; ORGANISM: homo sapiens  
US-10-214-811-10

Query Match 31.2% Score 5; DB 9; Length 211;  
Best Local Similarity 100.0%; Pred. No. 1.9e+02;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 10 RDPSP 14

DB 77 RDPSP 81

RESULT 25

US-09-780-016-10  
; Sequence 10, Application US/09780016  
; Patent No. US20020004591A1  
; GENERAL INFORMATION:  
; APPLICANT: Donoho, Gregory  
; APPLICANT: Scoville, John  
; APPLICANT: Turner, C. Alexander Jr.  
; APPLICANT: Friedlich, Glenn  
; APPLICANT: Abuin, Alejandro  
; APPLICANT: Zambrowicz, Brian  
; APPLICANT: Sands, Arthur T.  
; TITLE OF INVENTION: No. US20020004591A1el Human Proteases and  
; TITLE OF INVENTION: Polynucleotides Encoding the Same  
; FILE REFERENCE: LEX-0132-USA  
; CURRENT FILING DATE: US/09/780, 016  
; PRIOR FILING DATE: 2001-02-09  
; PRIOR APPLICATION NUMBER: US 60/181,294  
; PRIOR FILING DATE: 2000-02-11  
; NUMBER OF SEQ ID NOS: 27  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 10  
; LENGTH: 211  
; TYPE: PRT

; ORGANISM: homo sapiens

Query Match 31.2% Score 5; DB 10; Length 211;  
Best Local Similarity 100.0%; Pred. No. 1.9e+02;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 10 RDPSP 14

DB 77 RDPSP 81

RESULT 26

US-09-738-626-3636  
; Sequence 16, Application US/09738626  
; Publication No. US20020197605A1  
; GENERAL INFORMATION:  
; APPLICANT: NAKAGAWA, SATOSHI  
; APPLICANT: MIZOGUCHI, HIROSHI  
; APPLICANT: ANDO, SEIKO  
; APPLICANT: HAYASHI, MIKIRO  
; APPLICANT: OCHIAI, KEIKO  
; APPLICANT: YOKOI, HARUHIKO  
; APPLICANT: TATEISHI, NAOKO  
; APPLICANT: SENOH, AKIHIRO  
; APPLICANT: IKEDA, MASATO  
; APPLICANT: OZAKI, AKIO  
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES  
; FILE REFERENCE: 249-125  
; CURRENT FILING DATE: US/09/738, 626  
; PRIOR FILING DATE: 2000-12-18  
; PRIOR APPLICATION NUMBER: JP 99/377484  
; PRIOR FILING DATE: 1999-12-16  
; PRIOR APPLICATION NUMBER: JP 00/159162  
; PRIOR FILING DATE: 2000-04-07  
; PRIOR APPLICATION NUMBER: JP 00/280988  
; PRIOR FILING DATE: 2000-08-03  
; NUMBER OF SEQ ID NOS: 7059  
; SOFTWARE: PatentIn ver. 3.0  
; SEQ ID NO 3636  
; LENGTH: 218  
; TYPE: PRT  
; ORGANISM: Corynebacterium glutamicum  
US-09-738-626-3636

Query Match 31.2% Score 5; DB 9; Length 218;  
Best Local Similarity 100.0%; Pred. No. 2e+02;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VGYGR 5

DB 133 VGYGR 137

RESULT 27

US-10-214-811-16  
; Sequence 16, Application US/10214811  
; Publication No. US20030023062A1  
; GENERAL INFORMATION:  
; APPLICANT: Donoho, Gregory  
; APPLICANT: Scoville, John  
; APPLICANT: Turner, C. Alexander Jr.  
; APPLICANT: Friedlich, Glenn  
; APPLICANT: Abuin, Alejandro  
; APPLICANT: Zambrowicz, Brian  
; APPLICANT: Sands, Arthur T.  
; TITLE OF INVENTION: No. US20030023062A1el Human Proteases and  
; TITLE OF INVENTION: Polynucleotides Encoding the Same  
; FILE REFERENCE: LEX-0132-USA  
; CURRENT FILING DATE: US/10/214, 811  
; PRIOR FILING DATE: 2002-08-07  
; PRIOR APPLICATION NUMBER: US/09/780, 016  
; PRIOR FILING DATE: 2001-02-09

```

; PRIOR APPLICATION NUMBER: US 60/181,294
; PRIOR FILING DATE: 2000-02-11
; NUMBER OF SEQ ID NOS: 27
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 16
; LENGTH: 242
; TYPE: PRT
; ORGANISM: homo sapiens
US-10-214-811-16

```

```

Query Match      31.2%; Score 5; DB 9; Length 242;
Best Local Similarity 100.0%; Pred. No. 2.2e+02;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

Cy 10 RDPSS 14
Db 133 RDPSS 137

```

```

RESULT 28
US-09-780-016-16
; Sequence 16, Application US/09780016
; Patent No. US2002004591A1
; GENERAL INFORMATION:
; APPLICANT: Donoho, Gregory
; APPLICANT: Scoville, John
; APPLICANT: Turner, C. Alexander Jr.
; APPLICANT: Friedrich, Glenn
; APPLICANT: Abuhin, Alejandro
; APPLICANT: Zambrowicz, Brian
; APPLICANT: Sands, Arthur T.
; TITLE OF INVENTION: Polynucleotides Encoding the Same
; FILE REFERENCE: Lex-0132-USA
; CURRENT APPLICATION NUMBER: US/09/780,016
; CURRENT FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: US 60/181,294
; PRIOR FILING DATE: 2000-02-11
; NUMBER OF SEQ ID NOS: 27
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 16
; LENGTH: 242
; TYPE: PRT
; ORGANISM: homo sapiens
US-09-780-016-16

```

```

Query Match      31.2%; Score 5; DB 10; Length 242;
Best Local Similarity 100.0%; Pred. No. 2.2e+02;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

Cy 10 RDPSS 14
Db 133 RDPSS 137

```

```

RESULT 29
US-09-983-802-488
; Sequence 488, Application US/09983802
; Publication No. US20030022185A1
; GENERAL INFORMATION:
; APPLICANT: Fischer et al.
; TITLE OF INVENTION: 123 Human Secreted Proteins
; FILE REFERENCE: P2010P1
; CURRENT APPLICATION NUMBER: US/09/983,802
; CURRENT FILING DATE: 2001-10-25
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 09/227,357
; PRIOR FILING DATE: EARLIER FILING DATE: 1999-01-08
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: PCT/US98/13684
; PRIOR FILING DATE: EARLIER FILING DATE: 1998-07-07
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/051,926
; PRIOR FILING DATE: EARLIER FILING DATE: 1997-07-08
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/052,793
; PRIOR FILING DATE: EARLIER FILING DATE: 1997-07-08

```

```

; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/051,925
; PRIOR FILING DATE: EARLIER FILING DATE: 1997-07-08
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/051,929
; PRIOR FILING DATE: EARLIER FILING DATE: 1997-07-08
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/052,803
; PRIOR FILING DATE: EARLIER FILING DATE: 1997-07-08
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/052,732
; PRIOR FILING DATE: EARLIER FILING DATE: 1997-07-08
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/051,931
; PRIOR FILING DATE: EARLIER FILING DATE: 1997-07-08
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/051,932
; PRIOR FILING DATE: EARLIER FILING DATE: 1997-07-08
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/051,916
; PRIOR FILING DATE: EARLIER FILING DATE: 1997-07-08
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/051,930
; PRIOR FILING DATE: EARLIER FILING DATE: 1997-07-08
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/051,918
; PRIOR FILING DATE: EARLIER FILING DATE: 1997-07-08
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/051,920
; PRIOR FILING DATE: EARLIER FILING DATE: 1997-07-08
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/052,733
; PRIOR FILING DATE: EARLIER FILING DATE: 1997-07-08
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/052,795
; PRIOR FILING DATE: EARLIER FILING DATE: 1997-07-08
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/051,919
; PRIOR FILING DATE: EARLIER FILING DATE: 1997-07-08
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/051,928
; PRIOR FILING DATE: EARLIER FILING DATE: 1997-07-08
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/055,722
; PRIOR FILING DATE: EARLIER FILING DATE: 1997-08-18
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/055,723
; PRIOR FILING DATE: EARLIER FILING DATE: 1997-08-18
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/055,948
; PRIOR FILING DATE: EARLIER FILING DATE: 1997-08-18
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/055,949
; PRIOR FILING DATE: EARLIER FILING DATE: 1997-08-18
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/055,953
; PRIOR FILING DATE: EARLIER FILING DATE: 1997-08-18
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/055,950
; PRIOR FILING DATE: EARLIER FILING DATE: 1997-08-18
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/055,947
; PRIOR FILING DATE: EARLIER FILING DATE: 1997-08-18
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/055,964
; PRIOR FILING DATE: EARLIER FILING DATE: 1997-08-18
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/056,360
; PRIOR FILING DATE: EARLIER FILING DATE: 1997-08-18
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/055,684
; PRIOR FILING DATE: EARLIER FILING DATE: 1997-08-18
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/055,984
; PRIOR FILING DATE: EARLIER FILING DATE: 1997-08-18
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/055,954
; PRIOR FILING DATE: EARLIER FILING DATE: 1997-08-18
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/058,785
; PRIOR FILING DATE: EARLIER FILING DATE: 1997-09-12
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/058,664
; PRIOR FILING DATE: EARLIER FILING DATE: 1997-09-12
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/058,660
; PRIOR FILING DATE: EARLIER FILING DATE: 1997-09-12
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/058,661
; PRIOR FILING DATE: EARLIER FILING DATE: 1997-09-12
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 488
; LENGTH: 255
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-983-802-488

```

```

Query Match      31.2%; Score 5; DB 9; Length 255;
Best Local Similarity 100.0%; Pred. No. 2.3e+02;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Cy 10 RDPSS 14

```

Db 245 RDPsR 249

RESULT 30  
US-10-214-811-8

```
/ Sequence 8, Application US/10214811
/ Publication No. US20030023062A1
/ GENERAL INFORMATION:
/ APPLICANT: Donoho, Gregory
/ APPLICANT: Scoville, John
/ APPLICANT: Turner, C. Alexander Jr.
/ APPLICANT: Friedrich, Glenn
/ APPLICANT: Abuhin, Alejandro
/ APPLICANT: Zambrowicz, Brian
/ APPLICANT: Sands, Arthur T.
/ TITLE OF INVENTION: Polynucleotides Encoding the Same
/ FILE REFERENCE: LEX-0132-USA
/ CURRENT FILING DATE: 2002-08-07
/ PRIOR APPLICATION NUMBER: US/09/780,016
/ PRIOR FILING DATE: 2001-02-09
/ PRIOR APPLICATION NUMBER: US 60/181,294
/ PRIOR FILING DATE: 2000-02-11
/ NUMBER OF SEQ ID NOS: 27
/ SOFTWARE: FastSeq for Windows Version 4.0
/ SEQ ID NO 8
/ LENGTH: 265
/ TYPE: PRT
/ ORGANISM: homo sapiens
US-10-214-811-8
```

Query Match 31.2% Score 5; DB 9; Length 265;  
Best Local Similarity 100.0%; Pred. No. 2.3e+02;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 10 RDPsR 14  
Db 156 RDPsR 160

RESULT 31

```
US-09-780-016-8
/ Sequence 8, Application US/09780016
/ Patent No. US20020004591A1
/ GENERAL INFORMATION:
/ APPLICANT: Donoho, Gregory
/ APPLICANT: Scoville, John
/ APPLICANT: Turner, C. Alexander Jr.
/ APPLICANT: Friedrich, Glenn
/ APPLICANT: Abuhin, Alejandro
/ APPLICANT: Zambrowicz, Brian
/ APPLICANT: Sands, Arthur T.
/ TITLE OF INVENTION: No. US20020004591A1 Human Proteases and
/ TITLE OF INVENTION: Polynucleotides Encoding the Same
/ FILE REFERENCE: LEX-0132-USA
/ CURRENT APPLICATION NUMBER: US/09/780,016
/ CURRENT FILING DATE: 2001-02-09
/ PRIOR APPLICATION NUMBER: US 60/181,294
/ PRIOR FILING DATE: 2000-02-11
/ NUMBER OF SEQ ID NOS: 27
/ SOFTWARE: FastSeq for Windows Version 4.0
/ SEQ ID NO 8
/ LENGTH: 265
/ TYPE: PRT
/ ORGANISM: homo sapiens
US-09-780-016-8
```

Query Match 31.2% Score 5; DB 10; Length 265;  
Best Local Similarity 100.0%; Pred. No. 2.3e+02;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 10 RDPsR 14  
Db 156 RDPsR 160

RESULT 32

```
US-10-214-811-12
/ Sequence 12, Application US/10214811
/ Publication No. US20030023062A1
/ GENERAL INFORMATION:
/ APPLICANT: Donoho, Gregory
/ APPLICANT: Scoville, John
/ APPLICANT: Turner, C. Alexander Jr.
/ APPLICANT: Friedrich, Glenn
/ APPLICANT: Abuhin, Alejandro
/ APPLICANT: Zambrowicz, Brian
/ APPLICANT: Sands, Arthur T.
/ TITLE OF INVENTION: No. US20030023062A1 Human Proteases and
/ TITLE OF INVENTION: Polynucleotides Encoding the Same
/ FILE REFERENCE: LEX-0132-USA
/ CURRENT FILING DATE: 2002-08-07
/ PRIOR APPLICATION NUMBER: US/10/214,811
/ PRIOR FILING DATE: 2001-02-09
/ PRIOR APPLICATION NUMBER: US 60/181,294
/ PRIOR FILING DATE: 2000-02-11
/ NUMBER OF SEQ ID NOS: 27
/ SOFTWARE: FastSeq for Windows Version 4.0
/ SEQ ID NO 12
/ LENGTH: 267
/ TYPE: PRT
/ ORGANISM: homo sapiens
US-10-214-811-12
```

Query Match 31.2% Score 5; DB 9; Length 267;  
Best Local Similarity 100.0%; Pred. No. 2.3e+02;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 10 RDPsR 14  
Db 133 RDPsR 137

RESULT 33

```
US-09-780-016-12
/ Sequence 12, Application US/09780016
/ Patent No. US20020004591A1
/ GENERAL INFORMATION:
/ APPLICANT: Donoho, Gregory
/ APPLICANT: Scoville, John
/ APPLICANT: Turner, C. Alexander Jr.
/ APPLICANT: Friedrich, Glenn
/ APPLICANT: Abuhin, Alejandro
/ APPLICANT: Zambrowicz, Brian
/ APPLICANT: Sands, Arthur T.
/ TITLE OF INVENTION: No. US20020004591A1 Human Proteases and
/ TITLE OF INVENTION: Polynucleotides Encoding the Same
/ FILE REFERENCE: LEX-0132-USA
/ CURRENT APPLICATION NUMBER: US/09/780,016
/ CURRENT FILING DATE: 2001-02-09
/ PRIOR APPLICATION NUMBER: US 60/181,294
/ PRIOR FILING DATE: 2000-02-11
/ NUMBER OF SEQ ID NOS: 27
/ SOFTWARE: FastSeq for Windows Version 4.0
/ SEQ ID NO 12
/ LENGTH: 267
/ TYPE: PRT
/ ORGANISM: homo sapiens
US-09-780-016-12
```

Query Match 31.2% Score 5; DB 10; Length 267;  
Best Local Similarity 100.0%; Pred. No. 2.3e+02;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 10 RDPSR 14  
|||||  
Db 133 RDPSR 137

## RESULT 34

US-10-214-811-6  
; Sequence 6, Application US/10214811  
; Publication No. US20030023062A1  
; GENERAL INFORMATION:  
; APPLICANT: Donoho, Gregory  
; APPLICANT: Scoville, John  
; APPLICANT: Turner, C. Alexander Jr.  
; APPLICANT: Friedlich, Glenn  
; APPLICANT: Abuln, Alejandro  
; APPLICANT: Zambrowicz, Brian  
; APPLICANT: Sands, Arthur T.  
; TITLE OF INVENTION: Polynucleotides Encoding the Same  
; FILE REFERENCE: LEX-0132-USA  
; CURRENT APPLICATION NUMBER: US/10/214,811  
; PRIOR FILING DATE: 2002-08-07  
; PRIOR APPLICATION NUMBER: US/09/780,016  
; PRIOR FILING DATE: 2001-02-09  
; PRIOR APPLICATION NUMBER: US 60/181,294  
; PRIOR FILING DATE: 2000-02-11  
; NUMBER OF SEQ ID NOS: 27  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 6  
; LENGTH: 290  
; TYPE: PRT  
; ORGANISM: homo sapiens  
US-10-214-811-6

Query Match 31.2%; Score 5; DB 9; Length 290;  
Best Local Similarity 100.0%; Pred. No. 2.5e+02;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 10 RDPSR 14  
|||||  
Db 156 RDPSR 160

## RESULT 35

US-09-780-016-6  
; Sequence 6, Application US/09780016  
; Patent No. US20020004591A1  
; GENERAL INFORMATION:  
; APPLICANT: Donoho, Gregory  
; APPLICANT: Scoville, John  
; APPLICANT: Turner, C. Alexander Jr.  
; APPLICANT: Friedlich, Glenn  
; APPLICANT: Abuln, Alejandro  
; APPLICANT: Zambrowicz, Brian  
; APPLICANT: Sands, Arthur T.  
; TITLE OF INVENTION: No. US20020004591A1el Human Proteases and  
; FILE OF INVENTION: Polynucleotides Encoding the Same  
; FILE REFERENCE: LEX-0132-USA  
; CURRENT APPLICATION NUMBER: US/09/780,016  
; CURRENT FILING DATE: 2001-02-09  
; PRIOR APPLICATION NUMBER: US 60/181,294  
; PRIOR FILING DATE: 2000-02-11  
; NUMBER OF SEQ ID NOS: 27  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 6  
; LENGTH: 290  
; TYPE: PRT  
; ORGANISM: homo sapiens  
US-09-780-016-6

Query Match 31.2%; Score 5; DB 10; Length 290;  
Best Local Similarity 100.0%; Pred. No. 2.5e+02;

Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Qy 10 RDPSR 14  
|||||  
Db 156 RDPSR 160

## RESULT 36

US-09-510-332-151  
; Sequence 151, Application US/09510332  
; Publication No. US20030022278A1  
; GENERAL INFORMATION:  
; APPLICANT: Zuker, Charles S.  
; APPLICANT: Adler, Jon Elliot  
; APPLICANT: Ryba, Nick  
; APPLICANT: Mueller, Ken  
; APPLICANT: Hoon, Mark  
; APPLICANT: The Regents of the University of California  
; TITLE OF INVENTION: T2R, a No. US20030022278A1el Family of Taste Receptors  
; FILE REFERENCE: 02307E-098010US  
; CURRENT APPLICATION NUMBER: US/09/510,332  
; CURRENT FILING DATE: 2000-02-22  
; PRIOR APPLICATION NUMBER: US 09/393,634  
; PRIOR FILING DATE: 1999-09-10  
; NUMBER OF SEQ ID NOS: 172  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 151  
; LENGTH: 309  
; TYPE: PRT  
; ORGANISM: Mus sp.  
; FEATURE:  
; OTHER INFORMATION: mouse T2R24 (mGR24)  
US-09-510-332-151

Query Match 31.2%; Score 5; DB 9; Length 309;  
Best Local Similarity 100.0%; Pred. No. 2.7e+02;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 9 DRDPS 13  
|||||  
Db 217 DRDPS 221

## RESULT 37

US-09-738-626-3508  
; Sequence 3508, Application US/09738626  
; Publication No. US20020197605A1  
; GENERAL INFORMATION:  
; APPLICANT: NAKAGAWA, SATOSHI  
; APPLICANT: MIZOGUCHI, HIROSHI  
; APPLICANT: ANDO, SEIKO  
; APPLICANT: HAYASHI, MIKIRO  
; APPLICANT: OCHIAI, KEIKO  
; APPLICANT: YOKOI, HARUHIKO  
; APPLICANT: TATEISHI, NAOKO  
; APPLICANT: SENO, AKIHIRO  
; APPLICANT: IKEDA, MASATO  
; APPLICANT: OZAKI, AKIO  
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES  
; FILE REFERENCE: 249-125  
; CURRENT APPLICATION NUMBER: US/09/738,626  
; CURRENT FILING DATE: 2000-12-18  
; PRIOR APPLICATION NUMBER: JP 99/377484  
; PRIOR FILING DATE: 1999-12-16  
; PRIOR APPLICATION NUMBER: JP 00/159162  
; PRIOR FILING DATE: 2000-04-07  
; PRIOR APPLICATION NUMBER: JP 00/280988  
; PRIOR FILING DATE: 2000-08-03  
; NUMBER OF SEQ ID NOS: 7059  
; SOFTWARE: PatentIn ver. 3.0  
; SEQ ID NO 3508  
; LENGTH: 322  
; TYPE: PRT

; ORGANISM: Corynebacterium glutamicum  
US-09-738-626-3508

Query Match 31.2%; Score 5; DB 9; Length 322;  
Best Local Similarity 100.0%; Pred.No. 2.8e+02;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 VGYGR 5  
|||||  
DB 221 VGYGR 225

RESULT 38  
US-09-988-929A-1  
; Sequence 1, Application US/09988929A  
; Patent No. US20020150982A1  
; GENERAL INFORMATION:  
; APPLICANT: Brookhaven Science Associates  
; APPLICANT: Shanklin, John  
; APPLICANT: Cahoon, Edgar B  
; TITLE OF INVENTION: Mutant Patcy Acid Desaturase  
; FILE REFERENCE: BSA 02-01; CIP of 09/233,856, filed 01/19/99  
; CURRENT APPLICATION NUMBER: US/09/988,929A  
; CURRENT FILING DATE: 2001-11-19  
; PRIOR APPLICATION NUMBER: 09/233,856  
; PRIOR FILING DATE: 1999-01-19  
; NUMBER OF SEQ ID NOS: 1  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 1  
; LENGTH: 363  
; TYPE: PRT  
; ORGANISM: Ricinus communis  
US-09-988-929A-1

Query Match 31.2%; Score 5; DB 10; Length 363;  
Best Local Similarity 100.0%; Pred.No. 3.1e+02;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 4 GRDDN 8  
|||||  
DB 273 GRDDN 277

RESULT 39  
US-10-017-145-1  
; Sequence 1, Application US/10017145  
; Patent No. US20020151019A1  
; GENERAL INFORMATION:  
; APPLICANT: Brookhaven Science Associates  
; APPLICANT: Shanklin, John  
; TITLE OF INVENTION: Mutant Patcy Acid Desaturase and Methode for Directed Mutagenesis  
; FILE REFERENCE: CIP of 09/328,550 filed June 9, 1999; which was a CIP of 09/233,856  
; January 19, 1999  
; CURRENT APPLICATION NUMBER: US/10/017,145  
; CURRENT FILING DATE: 2001-12-14  
; PRIOR APPLICATION NUMBER: 09/328,550  
; PRIOR FILING DATE: 1999-06-09  
; NUMBER OF SEQ ID NOS: 13  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 1  
; LENGTH: 363  
; TYPE: PRT  
; ORGANISM: Ricinus communis  
; FEATURE:  
; NAME/KEY: misc\_feature  
; OTHER INFORMATION: ricinus communis delta 9 18:0 Acyl ACP Desaturase  
US-10-017-145-1

Query Match 31.2%; Score 5; DB 12; Length 363;  
Best Local Similarity 100.0%; Pred.No. 3.1e+02;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 4 GRDDN 8

|||||  
DB 273 GRDDN 277

RESULT 40  
US-09-775-879-16  
; Sequence 16, Application US/09775879  
; Patent No. US20020068822A1  
; GENERAL INFORMATION:  
; APPLICANT: Choe, Sunghwa  
; APPLICANT: Feldmann, Kenneth A  
; TITLE OF INVENTION: Dwf7 MUTANTS  
; FILE REFERENCE: 2225-0003  
; CURRENT APPLICATION NUMBER: US/09/775,879  
; CURRENT FILING DATE: 2001-02-02  
; PRIOR APPLICATION NUMBER: 60/179,901  
; PRIOR FILING DATE: 2000-02-02  
; NUMBER OF SEQ ID NOS: 25  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 16  
; LENGTH: 365  
; TYPE: PRT  
; ORGANISM: delta-7 sterol C-5 desaturase (Saccharomyces cerevisiae)  
US-09-775-879-16

Query Match 31.2%; Score 5; DB 10; Length 365;  
Best Local Similarity 100.0%; Pred.No. 3.1e+02;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 6 DDNDR 10  
|||||  
DB 349 DDNDR 353

Search completed: April 9, 2003, 13:27:30  
Job time : 21.9016 secs

















```

983 4 25.0 289 1 CHLL_MARPO P06267 marchantia
984 4 25.0 289 1 DHP5_SYNY3 P73248 synecocyst
985 4 25.0 289 1 GLA_IACIA Q95e02 lactococcus
986 4 25.0 289 1 P0R1_RHOBL P39767 rhodospaerum
987 4 25.0 289 1 P0R1_STRMU P95778 streptococcus
988 4 25.0 289 1 TP_CAVPO Q9J1u8 cavia porce
989 4 25.0 290 1 IMP4_YEAST P53941 saccharomyces
990 4 25.0 290 1 STR_APICA Q16932 aplysia cal
991 4 25.0 290 1 TRX2_MOUSE O08550 mus musculu
992 4 25.0 290 1 Y43A_MYCPN P75150 mycoplasma
993 4 25.0 290 1 Y873_METHA Q58283 methanococcus
994 4 25.0 290 1 YMA3_YEAST Q04257 saccharomyces
995 4 25.0 291 1 CORC_VIBCH Q9Kte3 vibrio chol
996 4 25.0 291 1 DH11_SAISC Q29608 salmella sci
997 4 25.0 291 1 DM43_DIDMR P82957 didelphis m
998 4 25.0 291 1 FLS_MATIN O04395 matthiola i
999 4 25.0 291 1 N1H1_AZOCH P26248 azotobacter
1000 4 25.0 291 1 Y802_YEAST P38282 saccharomyces

```

## ALIGNMENTS

```

RESULT 1
CERC_SCHEMA STANDARD; PRT; 264 AA.

```

```

ID CERC_SCHEMA STANDARD; PRT; 264 AA.
AC P12546;
DT 01-OCT-1989 (Rel. 12, Created)
DT 01-OCT-1989 (Rel. 12, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Cercarial protease precursor (EC 3.4.21.-) (Cercarial elastase).
OS Schistosoma mansoni (Blood fluke).
OC Eukaryota; Metazoa; Platyhelminthes; Trematoda; Digenea; Strigeidida;
OC Schistosomatidae; Schistosomatidae; Schistosoma.
OX NCBI_TaxID=6183;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=86330818; PubMed=3166457;
RA Newport G.R., McKerrow J.H., Hedstrom R., Pettit M., McGarrigle L.,
RA Barr P.J., Agabian N.;
RT "Cloning of the proteinase that facilitates infection by schistosome
parasites."
RL J. Biol. Chem. 263:13179-13184 (1988).
CC -1- FUNCTION: THIS PROTEASE CLEAVES ELASTIN AND THUS FACILITATES
PENETRATION OF SCHISTOSOME PARASITE LARVAE THROUGH ELASTIN-
RICH TISSUE OF THE HOST.
CC -1- ENZYME REGULATION: ACTIVATED BY AN AUTOCATALYTIC MECHANISM.
CC -1- SUBCELLULAR LOCATION: ACETABULAR (PENETRATION) GLANDS.
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
CC -----
DR EMBL; J03946; AAA29864.1; -
DR PIR; A28942; A28942.
DR HSSP; P20231; IAAO.
DR MEROPS; S01.144; -
DR InterPro; IPR001314; Chymotrypsin.
DR InterPro; IPR001254; Ser_protease_Try.
DR Pfam; PF00089; trypsin; 1.
DR PRINTS; PR00722; CHYMOTRYPSIN.
DR SMART; SM00020; Tryp_SPC; 1.
DR PROSITE; PS50240; TRYPSIN_DOM; 1.
DR PROSITE; PS00134; TRYPSIN_HIS; 1.
DR PROSITE; PS00135; TRYPSIN_SER; FALSE_NEG.
KM Hydrolase; Serine protease; Zymogen; Signal.
FT SIGNAL 1 19 POTENTIAL.
FT PROPEP 20 27

```

```

FT CHAIN 28 264 CERCARIAL PROTEASE. (BY SIMILARITY).
FT ACT_SITE 68 126 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE 126 126 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE 218 218 CHARGE RELAY SYSTEM (BY SIMILARITY).
SQ SEQUENCE 264 AA; 28545 MW; E2E5129A7C5D5010 CRC64;

```

```

Query Match 100.0%; Score 16; DB 1; Length 264;
Best Local Similarity 100.0%; Pred. No. 1.6e-10;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 1 VGYGRDNDNDPSPSRKN 16
Db 163 VGYGRDNDNDPSPSRKN 178

```

```

RESULT 2
YN48_ARCFU STANDARD; PRT; 112 AA.

```

```

ID YN48_ARCFU STANDARD; PRT; 112 AA.
AC O30321;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical protein AF2348.
GN AF2348.
OS Archaeoglobus fulgidus.
OC Archaea; Euryarchaeota; Archaeoglobales;
OC Archaeoglobaceae; Archaeoglobus.
OX NCBI_TaxID=2234;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=VC-16 / DSM 4304 / ATCC 49558;
STRAIN=96049343; PubMed=9389475;
RA Klink H.-P., Clayton R.A., Tomb J.-F., Peterson K.B.,
RA Ketchum K.A., Dodson R.J., Gwinn M., Hickey E.K., Peterson J.D.,
RA Richardson D.L., Kervilavage A.R., Graham D.E., Kyprides N.C.,
RA Fleischmann R.D., Quackenbush J., Lee N.H., Sutton G.G., Gill S.,
RA Kirkness E.F., Dougherty B.A., McKenney K., Adams W.D., Loftus B.,
RA Peterson S., Reich C.I., McNeil L.K., Badger J.H., Glodok A., Zhou L.,
RA Overbeek R., Gocayne J.D., Weidman J.F., McDonald L., Uterback T.,
RA Corton M.D., Spriggs T., Artlich P., Kaine B.P., Sykes S.M.,
RA Sadow P.W., D'Andrea K.P., Bowman C., Fujii C., Garland S.A.,
RA Mason T.M., Olsen G.J., Fraser C.M., Smith H.O., Woese C.R.,
RA Venter J.C.;
RT "The complete genome sequence of the hyperthermophilic, sulphate-
reducing archaeon Archaeoglobus fulgidus."
RL Nature 390:364-370 (1997).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AE001114; AAB91315.1; -
DR TIGR; AF2348; -
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 112 AA; 13012 MW; C9C6AC0ACD6AC730 CRC64;

```

```

Query Match 37.5%; Score 6; DB 1; Length 112;
Best Local Similarity 100.0%; Pred. No. 4.1;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 11 DPSRKN 16
Db 80 DPSRKN 85

```

```

RESULT 3
YN63_ARCFU STANDARD; PRT; 122 AA.
ID YN63_ARCFU STANDARD; PRT; 122 AA.
AC O28021;

```

DT 16-OCT-2001 (Rel. 40, Created)  
DT 16-OCT-2001 (Rel. 40, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE Hypothetical protein AF2263.  
GN AF2263.  
OS Archaeoglobus fulgidus.  
OC Archaea; Euryarchaeota; Archaeoglobi; Archaeoglobales;  
OC Archaeoglobaceae; Archaeoglobus.  
OX NCBI\_TaxID=2234;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=VC-16 / DSM 4304 / ATCC 49558;  
RX MEDLINE=98049343; PubMed=9389475;  
RA Klenk H.-P., Clayton R.A., Tomb J.-F., White O., Nelson K.E.,  
RA Kercham K.A., Dodson R.J., Gwinn M., Hickey E.K., Peterson J.D.,  
RA Richardson D.L., Kervagase A.R., Graham D.E., Kyriades N.C.,  
RA Fleischmann R.D., Quackenbush J., Lee N.H., Sutton G.G., Gill S.,  
RA Kinsess E.F., Dougherty B.A., McKenney K., Adams M.D., Loftus B.,  
RA Peterson S., Reich C.I., McNeil L.K., Badger J.H., Glodde A., Zhou L.,  
RA Overbeek R., Gocayne J.D., Weidman J.F., McDonald L., Uterback T.,  
RA Cotton M.D., Spriggs T., Artlich P., Kaine B.P., Sykes S.M.,  
RA Sadow P.W., D'Andrea K.P., Bowman C., Fujit C., Garland S.A.,  
RA Mason T.M., Olsen G.J., Fraser C.M., Smith H.O., Woese C.R.,  
RA Venter J.C.;  
RT "The complete genome sequence of the hyperthermophilic, sulphate-  
RL reducing archaeon Archaeoglobus fulgidus.";  
RL Nature 390:364-370(1997).  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL; AE000948; AAB8987.1; -.  
DR TIGR; AF2263; -.  
KM Hypothetical protein; Coiled coil; Complete proteome.  
FT DOMAIN 79 114 COILED COIL (POTENTIAL).  
SQ SEQUENCE 122 AA; 13824 MW; 31D01F2C8B0AA370 CRC64;  
Query Match 37.5%; Score 6; DB 1; Length 122;  
Best Local Similarity 100.0%; Pred. No. 4.5;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 2 GYGRDD 7  
Db 28 GYGRDD 33  
RESULT 4  
SR54\_MYCN STANDARD; PRT; 450 AA.  
ID SR54\_MYCN  
AC P75054;  
DT 01-NOV-1997 (Rel. 35, Created)  
DT 01-NOV-1997 (Rel. 35, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE Signal recognition particle protein (Fifty-Four homolog).  
GN F5F OR MPN061 OR MP093.  
OS Mycoplasma pneumoniae.  
OC Bacteria; Firmicutes; Mollicutes; Mycoplasmataceae; Mycoplasma.  
OX NCBI\_TaxID=2104;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=ATCC 29342 / M129;  
RX MEDLINE=97105885; PubMed=8948633;  
RA Himmelreich R., Hilbert H., Plagens H., Pirkel E., Li B.-C.,  
RA Herrmann R.;  
RT "Complete sequence analysis of the genome of the bacterium Mycoplasma  
RT pneumoniae.";  
RL Nucleic Acids Res. 24:4420-4449(1996).  
CC -I- FUNCTION: NECESSARY FOR EFFICIENT EXPORT OF EXTRA-CYTOPLASMIC

CC PROTEINS. BINDS TO THE SIGNAL SEQUENCE WHEN IT EMERGES FROM THE  
CC RIBOSOMES (BY SIMILARITY).  
CC -I- SUBUNIT: SIGNAL RECOGNITION PARTICLE CONSISTS OF A SMALL  
CC CYTOPLASMIC RNA (SC-RNA) MOLECULE AND PROTEIN F5F (BY SIMILARITY).  
CC -I- DOMAIN: THE PROTEIN HAS A TWO DOMAIN STRUCTURE: THE G-DOMAIN  
CC BINDS GTP; THE M-DOMAIN BINDS THE RNA AND ALSO BINDS THE SIGNAL  
CC SEQUENCE.  
CC -I- SIMILARITY: BELONGS TO THE SRP FAMILY OF GTP-BINDING PROTEINS.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL; AE000111; AAB95741.1; -.  
DR HSSP; O07347; IEPH.  
DR InterPro; IPR000897; SRP54.  
DR InterPro; IPR004125; SRP54\_SBP.  
DR InterPro; IPR004780; SRP\_sub.  
DR Pfam; PF00448; SRP54\_1.  
DR Pfam; PF02881; SRP54\_N; 1.  
DR Pfam; PF02978; SRP\_SBP; 1.  
DR ProDom; PD000819; SRP54\_1.  
DR TIGRFAMs; TIGR00959; 340501s01; 1.  
DR PROSITE; PS00300; SRP54\_1.  
KM Signal recognition particle; GTP-binding; RNA-binding;  
KM Complete proteome.  
FT DOMAIN 1 293 G-DOMAIN.  
FT DOMAIN 294 450 M-DOMAIN.  
FT NP\_BIND 106 113 GTP (BY SIMILARITY).  
FT NP\_BIND 188 192 GTP (BY SIMILARITY).  
FT NP\_BIND 246 249 GTP (BY SIMILARITY).  
SQ SEQUENCE 450 AA; 50141 MW; 7EF42ECCA4FE679 CRC64;  
Query Match 37.5%; Score 6; DB 1; Length 450;  
Best Local Similarity 100.0%; Pred. No. 14;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 10 RDP5RK 15  
Db 392 RDP5RK 397  
RESULT 5  
CATC\_MOUSE STANDARD; PRT; 462 AA.  
ID CATC\_MOUSE  
AC P97821; O08853;  
DT 15-JUL-1998 (Rel. 36, Created)  
DT 15-JUL-1998 (Rel. 36, Last sequence update)  
DT 15-JUN-2002 (Rel. 41, Last annotation update)  
DE Dipeptidyl-peptidase I precursor (EC 3.4.14.1) (DPP-I) (DPP1)  
DE (Cathepsin C) (Cathepsin J) (Dipeptidyl transferase).  
GN CTSC.  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
OX NCBI\_TaxID=10090;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=129/SvJ;  
RX MEDLINE=97256790; PubMed=909719;  
RA Pham C.T.N., Armstrong R.J., Zimonjic D.B., Popescu N.C., Payan D.G.,  
RA Ley T.J.;  
RT "Molecular cloning, chromosomal localization, and expression of  
RT murine dipeptidyl peptidase I.";  
RL J. Biol. Chem. 272:10695-10703(1997).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=BALB/c;  
RX MEDLINE=97276897; PubMed=9130590;

RA McGuire M.J., Lipsky P.E., Thiele D.L.;  
 RT "Cloning and characterization of the cDNA encoding mouse dipeptidyl  
 RT peptidase I (cathepsin C).";  
 RL Biochim. Biophys. Acta 1351:267-273(1997).  
 RN [3]  
 RP REVISIONS.  
 RC STRAIN=BAJB/c;  
 RA McGuire M.J., Lipsky P.E., Francisco N.M.C., Thiele D.L.;  
 RL Submitted (APR-2000) to the EMBL/GenBank/DBJ databases.  
 CC -1- FUNCTION: THIOLESTERASE. HAS DIPEPTIDYLPEPTIDASE ACTIVITY. CAN  
 CC DEGRADE GLUCAGON. PLAYS A ROLE IN THE GENERATION OF CYTOTOXIC  
 CC LYMPHOCYTE EFFECTOR FUNCTION.  
 CC -1- CATALYTIC ACTIVITY: Release of an N-terminal dipeptide, Xaa-Xbb-|-  
 CC xcc, except when Xaa is Arg or Lys, or Xbb or Xcc is Pro.  
 CC -1- SUBUNIT: DIMER OF AN ALPHA CHAIN AND A BETA CHAIN CROSS-LINKED  
 CC BY A DISULFIDE BOND.  
 CC -1- SUBCELLULAR LOCATION: Lysosomal.  
 CC -1- TISSUE SPECIFICITY: BROADLY DISTRIBUTED, BUT HIGHER LEVELS FOUND  
 CC IN LUNG, LIVER, KIDNEY AND SPLEEN. LOWER LEVELS FOUND IN TESTIS  
 CC AND BRAIN.  
 CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY C1.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC DR EMBL; U89269; AAB9457.1; -;  
 CC DR EMBL; U74683; AAB58400.3; -;  
 CC DR HSSP; O46427; 8PCH.  
 CC DR MEROPS; C01.070; -;  
 CC DR MGD; MGI:109553; Ctsc.  
 CC DR InterPro; IPR000668; peptidase\_C1.  
 CC DR InterPro; IPR000169; SHprot\_acsite.  
 CC DR Pfam; PF00112; Peptidase\_C1; 1.  
 CC DR PRINTS; PRO0705; PAPAIn\_1.  
 CC DR Prodom; PD000159; Peptidase\_C1; 1.  
 CC DR PROSITE; PS00139; THIOLESTERASE\_CYS; 1.  
 CC DR PROSITE; PS00639; THIOLESTERASE\_HIS; 1.  
 CC DR PROSITE; PS00640; THIOLESTERASE\_ASN; 1.  
 CC KW Hydroxylase; Thiol protease; Lysosome; Glycoprotein; Zymogen; Signal.  
 CC FT SIGNAL 1 24  
 CC FT PROPEP 25 230  
 CC FT CHAIN 231 393  
 CC FT CHAIN 231 393  
 CC FT ACT\_SITE 257 462  
 CC FT ACT\_SITE 257 462  
 CC FT ACT\_SITE 404 404  
 CC FT ACT\_SITE 426 426  
 CC FT CARBOHYD 29 29  
 CC FT CARBOHYD 53 53  
 CC FT CARBOHYD 275 275  
 CC SEQUENCE 462 AA; 52376 MW; 56574B3BD7F4710 CRC64;  
 Query Match 37.5%; Score 6; DB 1; Length 462;  
 Best Local Similarity 100.0%; Pred. No. 15;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 VGYGRD 6  
 DB 409 VGYGRD 414  
 RESULT 6  
 YN2 YEAST STANDARD; PRT; 1592 AA.  
 AC P53855;  
 DT 01-OCT-1996 (Rel. 34, Created)  
 DT 01-OCT-1996 (Rel. 34, Last sequence update)  
 DT 01-OCT-1996 (Rel. 34, Last annotation update)  
 DE Hypothetical 178.4 kDa protein in SLA2-ZWFI intergenic region.

GN YNL242W OR N1106.  
 OS Saccharomyces cerevisiae (Baker's yeast).  
 CC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
 CC Saccharomycetales; Saccharomycetaceae; Saccharomyces.  
 CC NCBI\_TaxID=4932;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=97051596; PubMed=8896273;  
 RA Pandolfo D., de Antoni A., Lanfranchi G., Valle G.;  
 RT "The DNA sequence of cosmid 14-5 from chromosome XIV reveals 21 open  
 RT reading frames including a novel gene encoding a globin-like  
 RT domain.";  
 RL Yeast 12:1071-1076(1996).  
 RN [2]  
 RP SEQUENCE OF 1-277 FROM N.A.  
 RA Satz J.E., Baladron V., del Rey F.;  
 RL Submitted (MAY-1996) to the EMBL/GenBank/DBJ databases.  
 RN [3]  
 RP SEQUENCE OF 278-1304 FROM N.A.  
 RA Garcia-Cantalejo J.M., Boskovic J., Jimenez A.;  
 RL Submitted (MAY-1996) to the EMBL/GenBank/DBJ databases.  
 CC -1- SIMILARITY: TO S. POMBE SPEC31E1.01C.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC DR EMBL; Z69381; CAA93356.1; -;  
 CC DR EMBL; Z71518; CAA96147.1; -;  
 CC DR SGD; S0005186; YNL242W.  
 CC KW Hypothetical protein.  
 CC SEQUENCE 1592 AA; 178412 MW; 54FDD56297FFP99 CRC64;  
 Query Match 37.5%; Score 6; DB 1; Length 1592;  
 Best Local Similarity 100.0%; Pred. No. 44;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 8 NDRDPS 13  
 DB 190 NDRDPS 195  
 RESULT 7  
 YFYD\_ECOLI STANDARD; PRT; 80 AA.  
 ID YFYD\_ECOLI  
 AC P76521;  
 DT 16-OCT-2001 (Rel. 40, Created)  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Hypothetical protein yfyd.  
 GN Escherichia coli.  
 OS Escherichia coli.  
 CC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;  
 CC Escherichia.  
 CC NCBI\_TaxID=562;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=X12 / MG1655;  
 RX MEDLINE=97426617; PubMed=9278503;  
 RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,  
 RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,  
 RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,  
 RA Mau B., Shao Y.;  
 RT "The complete genome sequence of Escherichia coli K-12.";  
 RL Science 277:1453-1474(1997).  
 CC -1- SUBCELLULAR LOCATION: Integral membrane protein (Potential).  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -



CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

DR EMBL; AE000326; AAC75436.1; -  
 DR Ecogene; EG14147; YfdY.  
 KM Hypothetical protein; Transmembrane; Complete proteome.  
 FT TRANSMEM 2 22 POTENTIAL.  
 FT TRANSMEM 32 52 POTENTIAL.  
 FT TRANSMEM 55 75 POTENTIAL.  
 SQ SEQUENCE 80 AA; 8874 MW; FB4195B3345DA5C CRC64;

Query Match 31.2%; Score 5; DB 1; Length 80;  
 Best Local Similarity 100.0%; Pred. No. 36;  
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VGYGR 5  
 |||||  
 Db 75 VGYGR 79

RESULT 8  
 Y123\_ECOLI STANDARD; PRT; 97 AA.  
 ID Y123\_ECOLI  
 AC P19778; P76918;  
 DT 01-FEB-1991 (Rel. 17, Created)  
 DT 01-FEB-1991 (Rel. 17, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Insertion element IS2 hypothetical 11.1 kDa protein (ORF3).  
 OS Escherichia coli.  
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;  
 OC Escherichia.  
 NCBI\_TaxID=562;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=K12;  
 RX MEDLINE=88137965; PubMed=2830172;  
 RA Ronecker H.J., Rak B.;  
 RT "Genetic organization of insertion element IS2 based on a revised  
 RT nucleotide sequence.";  
 RL Gene 59:291-296(1987).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=K12;  
 RX MEDLINE=97251357; PubMed=9097039;  
 RA Alba H., Baba T., Fujita K., Hayashi K., Inada T., Isono K.,  
 RA Itoh T., Kasai H., Kashimoto K., Kimura S., Kitakawa M.,  
 RA Kitagawa M., Makino K., Miki T., Mizobuchi K., Mori H., Mori T.,  
 RA Motomura K., Nakade S., Nakamura Y., Nishimoto H., Nishio Y.,  
 RA Oshima T., Saito N., Sempel G., Seki Y., Sivasubram S.,  
 RA Tagami H., Takeda U., Takemoto K., Takeuchi Y., Wada C.,  
 RA Yamamoto Y., Horiuchi T.;  
 RT "A 570-kb DNA sequence of the Escherichia coli K-12 genome  
 RT corresponding to the 28.0-40.1 min region on the linkage map.";  
 RL DNA Res. 3:363-377(1996).

CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC -----  
 DR EMBL; V00279; CAA23540.1; -  
 DR EMBL; D90778; BAA15012.1; -  
 DR EMBL; D90779; BAA15018.1; -  
 DR EMBL; D90838; BAA15821.1; -  
 DR EMBL; D90850; BAA16004.1; -  
 DR EMBL; D90851; BAA16012.1; -  
 DR EMBL; D90852; BAA22516.1; -  
 DR ALT\_INIT.

DR PIR; JQ0041; JQ0041.  
 KW Transposable element; Hypothetical protein.  
 SQ SEQUENCE 97 AA; 11147 MW; DBBE586EBBAA6EB CRC64;

Query Match 31.2%; Score 5; DB 1; Length 97;  
 Best Local Similarity 100.0%; Pred. No. 43;  
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VGYGR 5  
 |||||  
 Db 45 VGYGR 49

RESULT 9  
 YGF2\_YEAST STANDARD; PRT; 101 AA.  
 ID YGF2\_YEAST  
 AC P53175;  
 DT 01-OCT-1996 (Rel. 34, Created)  
 DT 01-OCT-1996 (Rel. 34, Last sequence update)  
 DT 01-NOV-1997 (Rel. 35, Last annotation update)  
 DE Hypothetical 12.1 kDa protein in Dupl-DUP2 intergenic region.  
 GN YGL052W.  
 OS Saccharomyces cerevisiae (Baker's yeast).  
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
 OC Saccharomycetales; Saccharomycetaceae; Saccharomyces.  
 NCBI\_TaxID=4932;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=S288C;  
 RX MEDLINE=97377993; PubMed=9234674;  
 RA Feuerhahn M., de Montigny J., Potier S., Souciet J.-L.;  
 RT "The characterization of two new clusters of duplicated genes  
 RT suggests a 'leugo' organization of the yeast Saccharomyces cerevisiae  
 RT chromosomes.";  
 RL Yeast 13:861-869(1997).

CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC -----  
 DR EMBL; Z72574; CAA96753.1; -  
 DR SGD; S0003020; YGL052W.  
 KW Hypothetical protein.  
 SQ SEQUENCE 101 AA; 12058 MW; DC74DA28F1B74162 CRC64;

Query Match 31.2%; Score 5; DB 1; Length 101;  
 Best Local Similarity 100.0%; Pred. No. 45;  
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 11 DPSRK 15  
 |||||  
 Db 86 DPSRK 90

RESULT 10  
 MIR2\_HUMAN STANDARD; PRT; 103 AA.  
 ID MIR2\_HUMAN  
 AC O9YH6;  
 DT 30-MAY-2000 (Rel. 39, Created)  
 DT 30-MAY-2000 (Rel. 39, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE Minimum potassium ion channel-related peptide 2 (MIRP2) (Mink-related  
 DE peptide 2).  
 GN KCNE3.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.  
 NCBI\_TaxID=9606;  
 RN [1]



```

DE D1 protein (Fragment) .
GN D1.
OS Onchocerca volvulus.
OC Eucaryota; Metazoa; Nematoda; Chromadorea; Spirurida; Filarioidea;
OC Onchocercidae; Onchocerca.
OX NCBI_TaxID=6282;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=97045813; PubMed=8890735;
RA Ertmann K.D., Gallin M.Y.;
RT "Onchocerca volvulus: identification of cDNAs encoding a putative
RT phosphatidyl-ethanolamine-binding protein and a putative partially
RT processed mRNA precursor.";
RL Gene 174:203-207(1996).
CC -1 SIMILARITY: BELONGS TO THE PHOSPHATIDYLETHANOLAMINE-BINDING
CC PROTEIN FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X87991; CAA61244.1; -.
DR HSSP; P13696; 1A44.
DR InterPro; IPR001858; PBP.
DR Pfam; PF01161; PBP; 1.
DR ProDom; PD004330; PBP; 1.
DR PROSITE; PS01220; PBP; 1.
FT NON TER 1
SQ SEQUENCE 152 AA; 16767 MW; E3B68C2E3E3B295D CRC64;

Query Match 31.2%; Score 5; DB 1; Length 152;
Best Local Similarity 100.0%; Pred. No. 65;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 12 PSRKN 16
   |||||
Db 57 PSRKN 61

RESULT 14
ILVH ECOLI STANDARD; PRT; 163 AA.
AC P00894; P78046; Q47637;
DT 21-JUL-1986 (Rel. 01, Created)
DT 01-AUG-1991 (Rel. 19, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Acetylactate synthase isozyme III small subunit (EC 4.1.3.18) (AHAS-
DE III) (Acetylhydroxy-acid synthase III small subunit) (ALS-III).
GN ILVH OR BBNP OR B0078.
OS Escherichia coli.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
OX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=83272971; PubMed=6308579;
RA Squires C.H., Defelice M., Devereux J., Calvo J.M.;
RT "Molecular structure of ilvH and its evolutionary relationship to
RT ilvG in Escherichia coli K12.";
RL Nucleic Acids Res. 11:5299-5313(1983).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=K12;
RX MEDLINE=92334977; PubMed=1630901;
RA Yura T., Mori H., Nagai H., Nagata T., Ishihama A., Fujita N.,
RA Isono K., Mizobuchi K., Nakata A.;
RT "Systematic sequencing of the Escherichia coli genome: analysis of
RT the 0-2.4 min region.";
RL Nucleic Acids Res. 20:3305-3308(1992).

```

```

RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=K12 / MG1655;
RX MEDLINE=97426617; PubMed=9278503;
RA Blatner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA Riley M., Collado-Valdes J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA Mau B., Shao Y.;
RT "The complete genome sequence of Escherichia coli K-12.";
RL Science 277:1453-1474(1997).
RN [4]
RP SEQUENCE OF 130-163 FROM N.A.
RC STRAIN=K12;
RX MEDLINE=91238714; PubMed=1851954;
RA Jahreis K., Postma P.W., Lengeler J.W.;
RT "Nucleotide sequence of the ilvH-fnr gene region of Escherichia coli
RT K12 and Salmonella typhimurium LT2.";
RL Mol. Gen. Genet. 226:332-336(1991).
RN [5]
RP SEQUENCE OF 130-163 FROM N.A.
RX MEDLINE=90330585; PubMed=2198273;
RA Leclerc G., Noel G., Drapeau G.R.;
RT "Molecular cloning, nucleotide sequence, and expression of shI, a new
RT gene in the 2-minute region of the genetic map of Escherichia coli.";
RL J. Bacteriol. 172:4696-4700(1990).
CC -1 CATALYTIC ACTIVITY: 2-acetylactate + CO(2) = 2 pyruvate.
CC -1 ENZYME REGULATION: SENSITIVE TO VALINE INHIBITION.
CC -1 PATHWAY: Valine and isoleucine biosynthesis; first step.
CC -1 SUBUNIT: DIMER OF LARGE AND SMALL CHAINS.
CC -1 SIMILARITY: BELONGS TO THE ACETOLACTATE SYNTHASE SMALL SUBUNIT
CC FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X01609; CAA25756.1; -.
DR EMBL; X55457; -; NOT ANNOTATED_CDS.
DR EMBL; X55034; CAA38855.1; -.
DR EMBL; D10483; BAA01346.1; -.
DR EMBL; AE000118; AAC73189.1; -.
DR PIR; A01114; YCEC3H.
DR PIR; M35034; AAA24627.1; -.
DR PIR; PS0152; PS0152.
DR PIR; S14386; S14386.
DR PIR; S40591; S40591.
DR SWISS-2DPAGE; P00894; COLI.
DR EcoGene; EG10499; ilvH.
DR InterPro; IPR002912; ACT.
DR InterPro; IPR004789; Acolac_sm.
DR Pfam; PF01842; ACT; 1.
DR TIGRPFAMs; TIGR00119; acolac_sm; 1.
DR ProDom; PD002844; Acolac_sm; 1.
RX Branched-chain amino acid biosynthesis; Lyase; Complete proteome.
FT CONFLICT 132 132 S -> D (IN REF. 3).
FT CONFLICT 159 163 DKIMR -> VK (IN REF. 1).
SQ SEQUENCE 163 AA; 17949 MW; 960C9A894A556078 CRC64;

Query Match 31.2%; Score 5; DB 1; Length 163;
Best Local Similarity 100.0%; Pred. No. 69;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 GYGRD 6
   |||||
Db 95 GYGRD 99

```

```

RESULT 15
RUVVC_THETN STANDARD; PRT; 165 AA.
AC OBRANA;
DT 15-JUN-2002 (Rel. 41, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DE CrossOver junction endonuclease ruvC (EC 3.1.22.4) (Holliday
  junction nuclease ruvC) (Holliday junction resolvase ruvC).
GN RUVVC OR TBE1178.
OS Thermobacter tengcongensis.
OC Bacteria; Firmicutes; Clostridia; Thermobacteriales;
OC Thermobacteriaceae; Thermobacter.
ON NCBI_TaxID=119072;
RX STRAIN=MB4 / JCM 11007;
RX MEDLINE=21992816; PubMed=11997336;
RA Bao Q., Tian Y., Li W., Xu Z., Xuan Z., Hu S., Dong W., Yang J.,
  Chen Y., Xue Y., Xu Y., Lai X., Huang L., Dong X., Ma Y., Ling L.,
  Tan H., Chen R., Wang J., Yu J., Yang H.;
  "A complete sequence of T. tengcongensis genome.";
  Genome Res. 12:689-700(2002).
RL
CC -!- FUNCTION: Nuclease that resolves Holliday junction intermediates
  in genetic recombination. Cleaves the cruciform structure in
  supercoiled DNA by nicking to strands with the same polarity at
  sites symmetrically opposed to the junction in the homologous arms
  and leaves a 5' terminal phosphate and a 3' terminal hydroxyl group
  (by similarity).
CC -!- CATALYTIC ACTIVITY: Endonucleolytic cleavage at a junction such as
  a reciprocal single-stranded crossover between two homologous DNA
  duplexes (Holliday junction).
CC -!- COFACTOR: Binds 1 magnesium ion per subunit (By similarity).
CC -!- SIMILARITY: BELONGS TO THE RUVVC FAMILY.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
  between the Swiss Institute of Bioinformatics and the EMBL outstation -
  the European Bioinformatics Institute. There are no restrictions on its
  use by non-profit institutions as long as its content is in no way
  modified and this statement is not removed. Usage by and for commercial
  entities requires a license agreement (See http://www.isb-sib.ch/announce/
  or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AE013080; AAM24409.1; -.
DR PROSITE; PS01321; RUVVC.1.
KW Hydrolyase; Nuclease; DNA repair; DNA recombination; Magnesium;
  Complete proteome.
FT METAL 7
FT METAL 67 MAGNESIUM (BY SIMILARITY).
FT METAL 140 MAGNESIUM (BY SIMILARITY).
FT METAL 143 MAGNESIUM (BY SIMILARITY).
SQ SEQUENCE 165 AA; 18267 MW; 9FA0074A37ADB9D7 CRC64;

Query Match 31.2%; Score 5; DB 1; Length 165;
Best Local Similarity 100.0%; Pred. No. 69;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Cy 1 VGYGR 5
Db 112 VGYGR 116

RESULT 16
CH18_DROGR STANDARD; PRT; 166 AA.
AC P24513;
DT 01-MAR-1992 (Rel. 21, Created)
DT 01-MAR-1992 (Rel. 21, Last sequence update)
DE 01-OCT-1996 (Rel. 34, Last annotation update)
DE Chorion protein S18.
GN CP18 OR S18.
OS Drosophila grimshawi (Fruit fly) (Idiomya grimshawi).
OC Eukaryota; Metazoa; Arthropoda; Mandibulata; Pancrustacea; Hexapoda;

```

```

CC Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
CC Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
ON NCBI_TaxID=7222;
RX SEQUENCE FROM N.A.
RX MEDLINE=91012583; PubMed=2213881;
RA Swammer C., Fennarjan M.G., Martinez-Cruzado J.C., Kafatos F.C.;
  "Evolution of the autosomal chorion cluster in Drosophila. III.
  Comparison of the s18 gene in evolutionarily distant species and
  heterospecific control of chorion gene amplification.";
  J. Mol. Biol. 215:225-235(1990).
RL
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
  between the Swiss Institute of Bioinformatics and the EMBL outstation -
  the European Bioinformatics Institute. There are no restrictions on its
  use by non-profit institutions as long as its content is in no way
  modified and this statement is not removed. Usage by and for commercial
  entities requires a license agreement (See http://www.isb-sib.ch/announce/
  or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X53422; CAA37504.1; -.
DR PIR; S13219; S13219.
DR FLYBASE; FBgn012319; Dgri\CP18.
KW Chorion.
SQ SEQUENCE 166 AA; 16810 MW; A710BEBE14F56FE CRC64;

Query Match 31.2%; Score 5; DB 1; Length 166;
Best Local Similarity 100.0%; Pred. No. 70;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Cy 1 VGYGR 5
Db 23 VGYGR 27

RESULT 17
RUVVC_RALSO STANDARD; PRT; 181 AA.
AC O8Y233;
DT 15-JUN-2002 (Rel. 41, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DE CrossOver junction endonuclease ruvC (EC 3.1.22.4) (Holliday
  junction nuclease ruvC) (Holliday junction resolvase ruvC).
GN RALSTONIA SOLANACEARUM (Pseudomonas solanacearum).
OS Bacteria; Proteobacteria; beta subdivision; Ralstonia group;
  Ralstonia.
ON NCBI_TaxID=305;
RX SEQUENCE FROM N.A.
RX STRAIN=GMT11000;
RX MEDLINE=21681879; PubMed=11823852;
RA Salanoubat M., Genin S., Artiguenave F., Guzy J., Mangenot S.,
  Ariat M., Billault A., Brotier P., Camus J.C., Cattolico L.,
  Chandler M., Choisan N., Claudel-Renard C., Cunne S., Demange N.,
  Gaspin C., Lavie M., Moisan A., Robert C., Saurin W., Schlex T.,
  Siguer P., Thebaud P., Whalen M., Winkler P., Levy M.,
  Weissenbach J., Boucher C.A.;
  "Genome sequence of the plant pathogen Ralstonia solanacearum.";
  Nature 415:497-502(2002).
RL
CC -!- FUNCTION: Nuclease that resolves Holliday junction intermediates
  in genetic recombination. Cleaves the cruciform structure in
  supercoiled DNA by nicking to strands with the same polarity at
  sites symmetrically opposed to the junction in the homologous arms
  and leaves a 5' terminal phosphate and a 3' terminal hydroxyl group
  (by similarity).
CC -!- CATALYTIC ACTIVITY: Endonucleolytic cleavage at a junction such as
  a reciprocal single-stranded crossover between two homologous DNA
  duplexes (Holliday junction).
CC -!- COFACTOR: Binds 1 magnesium ion per subunit (By similarity).
CC -!- SIMILARITY: BELONGS TO THE RUVVC FAMILY.
-----

```

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

DR EMBL; AL646059; CAD14031.1; -  
DR InterPro; IPR002176; RUVIC.  
DR Pfam; PF02075; RUVIC.1  
DR PRINTS; PR00696; RSOLVASERVIC.  
DR PRODOM; PD008333; RUVIC.  
DR TIGRFAMs; TIGR00228; RUVIC.1.  
DR PROSITE; PS01321; RUVIC; FALSE NEG.  
KM Hydroxylase; Nuclease; DNA repair; DNA recombination; Magnesium;  
KW Complete proteome.  
FT METAL 7  
FT METAL 67  
FT METAL 139  
FT METAL 142  
SQ SEQUENCE 181 AA; 19260 MW; 180F7E04EB305EAC CRC64;

Query Match 31.2%; Score 5; DB 1; Length 181;  
Best Local Similarity 100.0%; Pred. No. 75;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 VGYGR 5  
DB 112 VGYGR 116

## RESULT 18

ID 21KD\_ZYMMO STANDARD; PRT; 184 AA.  
AC P15256;  
DT 01-APR-1990 (Rel. 14, Created)  
DT 01-APR-1990 (Rel. 14, Last sequence update)  
DT 15-DEC-1998 (Rel. 37, Last annotation update)  
DE 20.9 kDa protein (ORF 2).  
OS Zymomonas mobilis.  
OG Plasmid pM2.  
OC Bacteria; Proteobacteria; alpha subdivision; Sphingomonadaceae;  
OC Zymomonas.  
OX NCBI\_TaxID=542;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=ATCC 10988 / ZM1;  
RA Misawa N., Nakamura K.;  
RT "The nucleotide sequence of the 2.7 kilobase pair plasmid of Zymomonas  
RL mobilis ATCC 10988".  
J. Biotechnol. 12:63-70(1989).

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

DR EMBL; X14438; CAA32610.1; -  
DR PIR; S06695; S06695.  
KM plasmid.  
SQ SEQUENCE 184 AA; 20962 MW; 83D50576B6F7802B CRC64;

Query Match 31.2%; Score 5; DB 1; Length 184;  
Best Local Similarity 100.0%; Pred. No. 77;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 4 GRDDN 8  
DB 36 GRDDN 40

RESULT 19  
Y045\_NPVAC STANDARD; PRT; 192 AA.  
ID Y045\_NPVAC  
AC P41450;  
DT 01-NOV-1995 (Rel. 32, Created)  
DT 01-NOV-1995 (Rel. 32, Last sequence update)  
DT 01-NOV-1995 (Rel. 32, Last annotation update)  
DE Hypothetical 22.7 kDa protein in GFA-P73 intergenic region.  
OS Autographa californica nuclear polyhedrosis virus (AcMNPV).  
OC Viruses; dsDNA viruses, no RNA stage; Baculoviridae;  
OC Nucleopolyhedrovirus.  
OX NCBI\_TaxID=46015;  
RN [1]  
RP SEQUENCE FROM N.A.

RC STRAIN=C6;  
RX MEDLINE=94303173; PubMed=8030224;  
RA Ayres M.D., Howard S.C., Kuzio J., Lopez-Ferber M., Possee R.D.;  
RT "The complete DNA sequence of Autographa californica nuclear  
RL polyhedrosis virus";  
Virology 202:586-605(1994).

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

DR EMBL; L22858; AAA6675.1; -  
KW Hypothetical protein.  
SQ SEQUENCE 192 AA; 22651 MW; B2CDD9ACC247E2A7 CRC64;

Query Match 31.2%; Score 5; DB 1; Length 192;  
Best Local Similarity 100.0%; Pred. No. 80;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 5 RDDND 9  
DB 50 RDDND 54

RESULT 20  
OY16\_ONCVO STANDARD; PRT; 197 AA.  
ID OY16\_ONCVO  
AC P31729;  
DT 01-JUL-1993 (Rel. 26, Created)  
DT 01-OCT-1996 (Rel. 34, Last sequence update)  
DT 01-OCT-1996 (Rel. 34, Last annotation update)  
DE OY-16 antigen precursor.  
GN OY16.  
OS Onchocerca volvulus.  
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Spirurida; Filarioidea;  
OC Onchocercidae; Onchocerca.  
OX NCBI\_TaxID=6282;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=90158691; PubMed=1689459;  
RA Lobos E., Altmann M., Mengod G., Weiss N., Rudin W., Karam M.;  
RT "Identification of an Onchocerca volvulus cDNA encoding a  
RT low-molecular-weight antigen uniquely recognized by onchocerciasis  
RT patient sera";  
Mol. Biochem. Parasitol. 39:135-146(1990).

RL MoI. Biochem. Parasitol. 39:135-146(1990).  
RN [2]  
RP CONCEPTUAL TRANSLATION.  
RA Bairoch A.;  
RL unpublished observations (SEP-1996).  
CC -1- TISSUE SPECIFICITY: HYPODERMIS, CUTICLE AND UTERUS.  
CC -1- SIMILARITY: BELONGS TO THE PHOSPHATIDYLETHANOLAMINE-BINDING  
CC PROTEIN FAMILY.

-1- CAUTION: THIS IS A CONCEPTUAL TRANSLATION; A FRAMESHIFT WAS

```
CC      INTRODUCED IN POSITION 135 TO PRODUCE THE CORRECT C-TERMINUS.
-----
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC      or send an email to license@isb-sib.ch).
-----
DR      EMBL; M27807; AAA29411.1; ALT_FRAME.
DR      PIR; A44947; A44947.
DR      HSSP; P30086; 1BD9.
DR      InterPro; IPR001858; PBP.
DR      Pfam; PF01161; PBP; 1.
DR      ProDom; PD004330; PBP; 1.
DR      PROSITE; PS01220; PBP; 1.
KM      Antigen; Signal; Glycoprotein.
FT      SIGNAL 1 16
FT      CHAIN 17 197
FT      CARBOHYD 56 56 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT      CARBOHYD 61 61 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT      CARBOHYD 119 119 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT      CARBOHYD 124 124 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ      SEQUENCE 197 AA; 21773 MW; 999898F20FDE854E3 CRC64;

Query Match 31.2%; Score 5; DB 1; Length 197;
Best Local Similarity 100.0%; Pred. No. 81;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy      12 PSRRN 16
        |||||
Db      102 PSRRN 106

RESULT 21
FLRE_BOVIN STANDARD; PRT; 205 AA.
AC      P52556;
DT      01-OCT-1996 (Rel. 34, Last sequence update)
DT      01-OCT-1996 (Rel. 34, Last sequence update)
DT      01-NOV-1997 (Rel. 35, Last annotation update)
DE      Flavin reductase (EC 1.6.99.1) (FR) (NADPH-dependent diaphorase)
DE      (NADPH-flavin reductase) (FLR) (Biliverdin reductase B) (EC 1.3.1.24)
DE      (BVR-B) (Biliverdin-IX beta-reductase) (Green heme binding protein)
DE      (GHRP).
GN      BLVRB.
OS      Bos taurus (Bovine).
OC      Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC      Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
OC      Bovidae; Bovinae; Bos.
OX      NCBI_TaxID=9913;
RN      [1]
RP      SEQUENCE FROM N.A., AND SEQUENCE OF 36-57 AND 117-126.
RC      TISSUE=Liver;
RX      MEDLINE=95023903; PubMed=7937764;
RA      Quandt K.S., Hultquist D.B.;
RT      "Flavin reductase: sequence of cDNA from bovine liver and tissue
RT      distribution.";
RL      Proc. Natl. Acad. Sci. U.S.A. 91:9322-9326(1994).
RN      [2]
RP      SEQUENCE OF 1-36; 120-136 AND 145-168.
RC      TISSUE=Erythrocyte;
RX      MEDLINE=91298967; PubMed=2069570;
RA      Quandt K.S., Xu F., Chen P., Hultquist D.E.;
RT      "Evidence that the protein components of bovine erythrocyte green
RT      heme binding protein and flavin reductase are identical.";
RL      Biochem. Biophys. Res. Commun. 178:315-321(1991).
CC      -!- FUNCTION: CATALYZES ELECTRON TRANSFER FROM REDUCED PYRIDINE
CC      NUCLEOTIDES TO FLAVINS AS WELL AS METHYLENE BLUE, PYRROQUINOLINE
CC      QUINONE, RIBOFLAVIN, OR METHEMOGLOBIN. POSSIBLE ROLE IN PROTECTING
CC      CELLS FROM OXIDATIVE DAMAGE OR IN REGULATING IRON METABOLISM. IN
CC      THE LIVER, CONVERTS BILIVERDIN TO BILIRUBIN.
```

```
CC      -!- CATALYTIC ACTIVITY: NADPH + acceptor = NADP(+) + reduced acceptor.
CC      -!- CATALYTIC ACTIVITY: Bilirubin + NAD(P)(+) = biliverdin + NAD(P)H.
CC      -!- SUBUNIT: MONOMER (BY SIMILARITY).
CC      -!- SUBCELLULAR LOCATION: Cytoplasmic (Potential).
CC      -!- TISSUE SPECIFICITY: AT LEAST EXPRESSED IN THE LIVER AND
CC      ERYTHROCYTE.
CC      -!- MASS SPECTROMETRY: MW=21994; METHOD=Electrospray.
-----
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC      or send an email to license@isb-sib.ch).
-----
DR      EMBL; L35045; AAC37323.1; -.
KM      Oxidoreductase; NADP; Multifunctional enzyme.
FT      INIT MET 4 35
FT      NP BIND 0
SQ      SEQUENCE 205 AA; 22001 MW; E5E8FE29C190404 CRC64;

Query Match 31.2%; Score 5; DB 1; Length 205;
Best Local Similarity 100.0%; Pred. No. 84;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy      10 RDPSP 14
        |||||
Db      34 RDPSP 38

RESULT 22
ID_YSY2_SYP7 STANDARD; PRT; 208 AA.
AC      P37372;
DT      01-OCT-1994 (Rel. 30, Created)
DT      01-OCT-1994 (Rel. 30, Last sequence update)
DT      16-OCT-2001 (Rel. 40, Last annotation update)
DE      Hypothetical 23.0 kDa protein in SYNA 5' region (ORF 2).
DE      Synecococcus sp. (strain PCC 7942) (Anacystis nidulans R2).
OS      Bacteria; Cyanobacteria; Chroococcales; Synecococcus.
OX      NCBI_TaxID=1140;
RN      [1]
RP      SEQUENCE FROM N.A.
RX      MEDLINE=95023970; PubMed=7937823;
RA      Phung L.T., Ajlani G., Haselkorn R.;
RT      "p-type ATPase from the cyanobacterium Synecococcus 7942 related to
RT      the human Menkes and Wilson disease gene products.";
RL      Proc. Natl. Acad. Sci. U.S.A. 91:9651-9654(1994).
CC      -!-
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC      or send an email to license@isb-sib.ch).
-----
DR      EMBL; U04356; AAB82019.1; -.
KM      Hypothetical protein.
SQ      SEQUENCE 208 AA; 23031 MW; 20411E5FB3408DAE CRC64;

Query Match 31.2%; Score 5; DB 1; Length 208;
Best Local Similarity 100.0%; Pred. No. 85;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy      8 NDRDP 12
        |||||
Db      51 NDRDP 55

RESULT 23
KTHY_IACIA
```

```
ID KTHY_IACLA STANDARD; PRT; 211 AA.
AC Q9CIG4;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DE Thymidylate kinase (EC 2.7.4.9) (dTMP Kinase).
GN TMK OR LT0398.
OS Lactococcus lactis (subsp. lactis) (Streptococcus lactis).
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae; Lactococcus.
NCBI_TaxID=1160;
RN
RP SEQUENCE FROM N.A.
RC STRAIN=111403;
RX MEDLINE=21235186; PubMed=11337471;
RA Bolotin A., Winkler P., Manger S., Jallion O., Malarme K.,
RT Weissenbach J., Ehrlich S.D., Sorokin A.;
RT "The complete genome sequence of the lactic acid bacterium Lactococcus
RT lactis sep. lactis 111403."
RL Genome Res. 11:731-753(2001).
CC -1- FUNCTION: PHOSPHORYLATION OF DTPP TO FORM DTPP IN BOTH DE
CC NOVO AND SALVAGE PATHWAYS OF DTPP SYNTHESIS (BY SIMILARITY).
CC -1- CATALYTIC ACTIVITY: ATP + thymidine 5'-phosphate = ADP + thymidine
CC 5'-diphosphate.
CC -1- SIMILARITY: BELONGS TO THE THYMIDYLATE KINASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AE006276; AAK04496.1; -.
DR HSSP; P37345; STMP.
DR InterPro; IPR000062; Thymidylate_kin.
DR Pfam; PF02223; Thymidylate_kin; 1.
DR TIGRPFAMS; TIGR00041; dTMP_Kinase; 1.
DR PROSITE; PS01331; THYMIDYLATE_KINASE; 1.
KW Transferase; Kinase; Nucleotide biosynthesis; ATP-binding;
KM Complete proteome.
FT NP_BIND 10 17 ATP (POTENTIAL).
FT SEQUENCE 211 AA; 23999 MW; 1C1F2C5ED982A55 CRC64;
SQ
Query March 31.2%; Score 5; DB 1; Length 211;
Best Local Similarity 100.0%; Pred. No. 86;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 2 GYGRD 6
DB 106 GYGRD 110
RESULT 24
GT26_FASHE STANDARD; PRT; 217 AA.
AC P31012;
DT 01-APR-1993 (Rel. 25, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE Glutathione S-transferase 26 kDa 51 (EC 2.5.1.18) (GST51) (FHS1)
DE (GST class-alpha).
OS Fasciola hepatica (liver fluke).
OC Eukaryota; Metazoa; Platyhelminthes; Trematoda; Digenea;
OC Echinostomidae; Echinostomata; Fasciolidae; Fasciolidae; Fasciola.
NCBI_TaxID=6192;
RN
RP SEQUENCE FROM N.A.
RX MEDLINE=92155306; PubMed=1740183;
RA Panaccio M., Wilson L.R., Cramer S.L., Wiffels G.L., Spithill T.W.;
RT "Molecular characterization of cDNA sequences encoding glutathione S-
RT transferases of Fasciola hepatica."
RT Exp. Parasitol. 74:232-237(1992).
```

```
RN
RP ERRATUM.
RX MEDLINE=94039664; PubMed=8224094;
RA Panaccio M., Wilson L.R., Cramer S.L., Wiffels G.L., Spithill T.W.;
RL Exp. Parasitol. 77:385-385(1993).
RN
RN [3]
RX SEQUENCE OF 9-217 FROM N.A.
RX MEDLINE=93228188; PubMed=7682383;
RA Muro A., Rodriguez-Molina J.R., Hillyer G.V.;
RT "Sequence analysis of a Fasciola hepatica glutathione S-transferase
RT cDNA clone."
RT Am. J. Trop. Med. Hyg. 48:457-463(1993).
CC -1- FUNCTION: CONJUGATION OF REDUCED GLUTATHIONE TO A WIDE NUMBER
CC OF EXOGENOUS AND ENDOGENOUS HYDROPHOBIC ELECTROPHILES.
CC -1- FUNCTION: GST ISOENZYMES APPEAR TO PLAY A CENTRAL ROLE IN THE
CC PARASITE DETOXIFICATION SYSTEM. OTHER FUNCTIONS ARE ALSO
CC SUSPECTED INCLUDING A ROLE IN INCREASING THE SOLUBILITY OF
CC HEMATIN IN THE PARASITE GUT.
CC -1- CATALYTIC ACTIVITY: RX + glutathione = HX + R-S-glutathione.
CC -1- SUBUNIT: HOMODIMER.
CC -1- SUBCELLULAR LOCATION: Cytoplasmic.
CC -1- SIMILARITY: BELONGS TO THE GST SUPERFAMILY. ALPHA FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M77682; AAA29141.1; -.
DR HSSP; P31670; IPHE.
DR InterPro; IPR004046; GST_Cterm.
DR InterPro; IPR004045; GST_Nterm.
DR Pfam; PF00043; GST_C; 1.
DR Pfam; PF02798; GST_N; 1.
KW Transferase; Antigen; Multigene family.
FT INIT_MET 0 BY SIMILARITY.
FT CONFLICT 83 83 T -> S (IN REF. 3).
FT CONFLICT 134 134 N -> D (IN REF. 3).
FT CONFLICT 146 146 P -> T (IN REF. 3).
FT SEQUENCE 217 AA; 25242 MW; EFCABF7454BDC26 CRC64;
SQ
Query March 31.2%; Score 5; DB 1; Length 217;
Best Local Similarity 100.0%; Pred. No. 89;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 3 YGRDD 7
DB 32 YGRDD 36
RESULT 25
GT27_FASHE STANDARD; PRT; 217 AA.
AC P31670;
DT 01-JUL-1993 (Rel. 26, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE Glutathione S-transferase 26 kDa 47 (EC 2.5.1.18) (GST47) (FH47)
DE (GST class-alpha).
OS Fasciola hepatica (liver fluke).
OC Eukaryota; Metazoa; Platyhelminthes; Trematoda; Digenea;
OC Echinostomidae; Echinostomata; Fasciolidae; Fasciolidae; Fasciola.
NCBI_TaxID=6192;
RN
RP SEQUENCE FROM N.A.
RX MEDLINE=92155306; PubMed=1740183;
RA Panaccio M., Wilson L.R., Cramer S.L., Wiffels G.L., Spithill T.W.;
RT "Molecular characterization of cDNA sequences encoding glutathione S-
RT transferases of Fasciola hepatica."
RT Exp. Parasitol. 74:232-237(1992).
```

```
RN [2]
RP ERRATUM.
RX MEDLINE=94039664; PubMed=8224094;
RA Panaccio M., Wilson L.R., Cramer S.L., Wijffels G.L., Spithill T.W.;
RL Exp. Parasitol. 77:385-385(1993).
RN [3]
RP SEQUENCE OF 7-105 FROM N.A.
RA Cramer S.;
RL Patent number WO9008819, 09-AUG-1990.
RN [4]
RP X-RAY CRYSTALLOGRAPHY (3 ANGSTROMS).
RX MEDLINE=96035725; PubMed=9367777;
RA Rossjohn J., Fell S.C., Wilce M.C.J., Sexton J.L., Spithill T.W.,
RA Parker M.W.;
RT "Crystallization, structural determination and analysis of a novel
RT parasite vaccine candidate: Fasciola hepatica glutathione
RT S-transferase.";
RL J. Mol. Biol. 273:857-872(1997).
CC -1- FUNCTION: CONJUGATION OF REDUCED GLUTATHIONE TO A WIDE NUMBER
CC OF EXOGENOUS AND ENDOGENOUS HYDROPHOBIC ELECTROPHILES.
CC -1- FUNCTION: GST ISOENZYMES APPEAR TO PLAY A CENTRAL ROLE IN THE
CC PARASITE DETOXIFICATION SYSTEM. OTHER FUNCTIONS ARE ALSO
CC SUSPECTED INCLUDING A ROLE IN INCREASING THE SOLUBILITY OF
CC HAEMATIN IN THE PARASITE GUT.
CC -1- CATALYTIC ACTIVITY: RX + glutathione = HX + R-S-glutathione.
CC -1- SUBUNIT: HOMODIMER.
CC -1- SIMILARITY: BELONGS TO THE GST SUPERFAMILY. ALPHA FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M77681; AAA29140.1; -.
DR EMBL; A00996; CA00121.1; -.
DR PDB; 1FHE; 29-JUL-98.
DR InterPro; IPR004046; GST_Cterm.
DR InterPro; IPR004045; GST_Nterm.
DR Pfam; PF00043; GST_N; 1.
DR Pfam; PF02798; GST_N; 1.
DR Transferrase; Antigen; Multigene family; 3D-structure.
KW INIT MET
FT INIT MET 0
FT CONFLICT 65 65 T->I (IN REF. 3).
FT CONFLICT 102 105 RIGF->FEEL (IN REF. 3).
SQ SEQUENCE 217 AA; 25281 MW; 0FB8BFE63029E03 CRC64;

Query Match 31.2%; Score 5; DB 1; Length 217;
Best Local Similarity 100.0%; Pred. No. 89;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 YGRDD 7
AC Q03489;
DB 32 YGRDD 36

RESULT 26
AGL9_PETHY STANDARD; PRT; 241 AA.
ID AGL9_PETHY
AC Q03489;
DT 01-FEB-1996 (Rel. 33, Last Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Agamous-like MADS box protein AGL9 homolog (Floral homeotic protein
DE FBP2) (Floral binding protein 2).
GN FBP2.
OS Petunia hybrida (Petunia).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
OC Asteridae; euasterids I; Solanales; Solanaceae; Petunia.
```

```
OX NCBI_TaxID=4102;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=93005737; PubMed=1356537;
RA Angenot G.C., Buscher M., Franken J., Mol J.N.M., Van Tunen A.J.;
RT "Differential expression of two MADS box genes in wild-type and
RT mutant petunia flowers.";
RL Plant Cell 4:983-993(1992).
RN [2]
RP REVISIONS TO C-TERMINUS.
RA Angenot G.C.;
RL Submitted (DEC-1992) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: PROBABLE TRANSCRIPTION FACTOR.
CC -1- SUBCELLULAR LOCATION: Nuclear.
CC -1- SIMILARITY: BELONGS TO THE MADS DOMAIN FAMILY OF TRANSCRIPTION
CC FACTORS.
CC -1- SIMILARITY: CONTAINS 1 K-BOX DIMERIZATION DOMAIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M91666; AAA86854.1; -.
DR PIR; J01690; J01690.
DR HSP; P11746; 1NMW.
DR TRANSFAC; T03093; -.
DR InterPro; IPR002487; TF_Knox.
DR InterPro; IPR002100; TF_MADSbox.
DR Pfam; PF00319; SRF-TF; 1.
DR Pfam; PF01486; K-box; 1.
DR PRINTS; PR00404; MADSDOMAIN.
DR SMART; SM00432; MADS_1.
DR PROSITE; PS00350; MADS_BOX_1; 1.
DR PROSITE; PS50066; MADS_BOX_2; 1.
DR Transcription regulation; DNA-binding; Nuclear protein.
KW DOMAIN
FT DOMAIN 3 57 MADS.
FT FT 98 170 K-BOX.
SQ SEQUENCE 241 AA; 27562 MW; B575BE9468D10BB CRC64;

Query Match 31.2%; Score 5; DB 1; Length 241;
Best Local Similarity 100.0%; Pred. No. 97;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VGYGR 5
AC Q38694;
DB 190 VGYGR 194

RESULT 27
AGL9_ARADE STANDARD; PRT; 250 AA.
ID AGL9_ARADE
AC Q38694;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Agamous-like MADS box protein AGL9 homolog (OM1).
OS Arabidopsis thaliana (Arabidopsis).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Asparagales; Orchidaceae;
OC Epidendroideae; higher Epidendroideae; Vandaceae; Acoridinae;
OC Arachnis x Vanda.
OX NCBI_TaxID=29714;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=94072738; PubMed=8251643;
RA Lu Z.X., Wu M., Loh C.S., Yeong C.Y., Goh C.J.;
RT "Nucleotide sequence of a flower-specific MADS box cDNA clone from
RT Plant Mol. Biol. 23:901-904(1993).
```



```

CC -1- FUNCTION: PROBABLE TRANSCRIPTION FACTOR ACTIVE IN INFLORESCENCE
CC DEVELOPMENT AND FLORAL ORGANORENSIS (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: Nuclear (By similarity).
CC -1- TISSUE SPECIFICITY: EXPRESSED IN PETALS AND WEAKLY IN SEPALs BUT
CC NOT IN THE COLUMN (GYNOSTEMIUM).
CC -1- DEVELOPMENTAL STAGE: EXPRESSED IN MATURE FLOWERS AND NOT IN YOUNG
CC DEVELOPING INFLORESCENCES OR YOUNG FLORAL BUDS.
CC -1- SIMILARITY: BELONGS TO THE MADS DOMAIN FAMILY OF TRANSCRIPTION
CC FACTORS.
CC -1- SIMILARITY: CONTAINS 1 K-BOX DIMERIZATION DOMAIN.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X69107; CAA48859.1; -.
DR HSSP; P11746; IMM.
DR TRNSPAC; T03114; -.
DR InterPro; IPR002487; TF_Kbox.
DR InterPro; IPR002100; TF_MADSbox.
DR Pfam; PF00319; SRF-TF; 1.
DR Pfam; PF01486; K-box; 1.
DR PRINTS; PRO0404; MADSDOMAIN.
DR SMART; SM00432; MADS; 1.
DR PROSITE; PS00350; MADS_BOX_1; 1.
DR PROSITE; PS50066; MADS_BOX_2; 1.
DR Transcription regulation; DNA-binding; Nuclear protein.
FT DOMAIN 3 57 MADS.
FT DOMAIN 96 168 K-BOX.
SQ SEQUENCE 250 AA; 28770 MW; 01BF9ADADC499C41 CRC64;

Query Match 31.2%; Score 5; DB 1; Length 250;
Best Local Similarity 100.0%; Pred. No. 1e+02;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VGYGR 5
DB 191 VGYGR 195

RESULT 28
TPIS_SCHEMA STANDARD; PRT; 253 AA.
AC P48501;
DT 01-FEB-1996 (Rel. 33, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 01-FEB-1996 (Rel. 33, Last annotation update)
DE Ttiosephosphate isomerase (EC 5.3.1.1) (TIM).
GN TPI.
OS Schistosoma mansoni (Blood fluke).
OC Eukaryota; Metazoa; Platyhelminthes; Trematoda; Digenea; Strigoidida;
OC Schistosomatidae; Schistosomatidae; Schistosoma.
OX NCBI_TaxID=6183;
[1]
RP SEQUENCE FROM N.A.
RX MEDLINE=92179278; PubMed=1542681;
RA Shoemaker C., Gross A., Gebremichael A., Harn D.;
RT "cDNA cloning and functional expression of the Schistosoma mansoni
RT protective antigen triose-phosphate isomerase.";
RL Proc. Natl. Acad. Sci. U.S.A. 89:1842-1846(1992).
[2]
RP SEQUENCE FROM N.A.
RX STRAIN=Puerto Rican;
RX MEDLINE=93341572; PubMed=8341322;
RA Dos Reis M.G., Davis R.E., Singh H., Skelly P.J., Shoemaker C.B.;
RT "Characterization of the Schistosoma mansoni gene encoding the
RT glycolytic enzyme, triosephosphate isomerase.";
RL Mol. Biochem. Parasitol. 59:235-242(1993).
[3]

```

```

RP SEQUENCE OF 87-98; 182-192 AND 210-222.
RX MEDLINE=92105755; PubMed=1729373;
RA Harn D.A., Gu W., Oligino L.D., Mitsuyama M., Gebremichael A.,
RA Richter D.;
RT "A protective monoclonal antibody specifically recognizes and alters
RT the catalytic activity of schistosome triose-phosphate isomerase.";
RL J. Immunol. 148:562-567(1992).
CC -1- FUNCTION: ANTIGEN TO THE HOST M.1 MONOCLONAL ANTIBODY.
CC -1- CATALYTIC ACTIVITY: D-glyceraldehyde 3-phosphate = glyceralone
CC phosphate.
CC -1- PATHWAY: PLAYS AN IMPORTANT ROLE IN SEVERAL METABOLIC PATHWAYS.
CC -1- SUBUNIT: HOMODIMER.
CC -1- SUBCELLULAR LOCATION: Cytoplasmic (probable).
CC -1- SIMILARITY: BELONGS TO THE TRIOSEPHOSPHATE ISOMERASE FAMILY.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M83294; AAA29941.1; -.
DR EMBL; L07286; AAA29919.1; -.
DR EMBL; L06636; AAA29919.1; JOINED.
DR EMBL; L07283; AAA29919.1; JOINED.
DR EMBL; L07284; AAA29919.1; JOINED.
DR EMBL; L07285; AAA29919.1; JOINED.
DR HSSP; P00940; ITPH.
DR InterPro; IPR000652; Triophos_ismrse.
DR Pfam; PF00121; TIM; 1.
DR ProDom; PD001005; Triophos_ismrse; 1.
DR TIGRfams; TIGR00419; tim; 1.
DR PROSITE; PS00171; TIM; 1.
KW Isomerase; Glycolysis; Gluconeogenesis; Fatty acid biosynthesis;
KW Pentose shunt.
FT ACT_SITE 96 96 BY SIMILARITY.
FT ACT_SITE 169 169 BY SIMILARITY.
SQ SEQUENCE 253 AA; 28122 MW; 883459137F6C9847 CRC64;

Query Match 31.2%; Score 5; DB 1; Length 253;
Best Local Similarity 100.0%; Pred. No. 1e+02;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 5 RDDND 9
DB 19 RDDND 23

RESULT 29
Y579_METTH STANDARD; PRT; 266 AA.
AC O26679;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Putative aldolase MTH579 (EC 4.2.1.-).
GN MTH579.
OS Methanobacterium thermoautotrophicum.
OC Archaea; Euryarchaeota; Methanobacteria; Methanobacteriales;
OC Methanobacteriaceae; Methanobacter.
OX NCBI_TaxID=187420;
[1]
RP SEQUENCE FROM N.A.
RX STRAIN=Delta H;
RX MEDLINE=98037514; PubMed=9371463;
RA Smith D.R., Doucette-Stamm L.A., Delonghery C., Lee H.-M., Dubois J.,
RA Aldredge T., Bashirzadeh R., Blakely D., Cook R., Gilbert K.,
RA Harrison D., Hoang L., Keagle P., Lumm W., Pochler B., Qiu D.,
RA Spadafora R., Viscare R., Wang Y., Wierzbowski J., Gibson R.,
RA Jiwani N., Caruso A., Bush D., Safer H., Patwell D., Prabhakar S.,
RA McDougall S., Shimer G., Goyal A., Pietrowski S., Church G.M.,

```

RA Daniels C.J., Mao J.-I., Rice P., Noelling J., Reeve J.N.;  
RT "Complete genome sequence of Methanobacterium thermoautotrophicum  
RL J. Bacteriol. 179:7135-7155 (1997).  
CC -1- SIMILARITY: BELONGS TO THE DEOC/FBAB FAMILY OF ALDOLASES.  
-----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
-----  
DR EMBL; A000840; AAB85085.1; -.  
DR InterPro; IPR002915; Deoc.  
DR Pfam; PF01791; Deoc; 1.  
KM Hypothetical protein; Lyase; Schiff base; Complete proteome.  
FT BINDING 176 176 SCHIFF-BASE (BY SIMILARITY).  
SQ SEQUENCE 266 AA; 28013 MW; 9381A9AAC28EB2C CRC64;  
  
Query March 31.2%; Score 5; DB 1; Length 266;  
Best Local Similarity 100.0%; Pred. No. 1.1e+02;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
QY 2 GYGRD 6  
Db 67 GYGRD 71  
  
RESULT 30  
VG18\_BPP22 STANDARD; PRT; 271 AA.  
ID P03687;  
DT 21-JUL-1986 (Rel. 01, Created)  
DT 21-JUL-1986 (Rel. 01, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE DNA replication protein gp18.  
DN 18.  
OS Bacteriophage P22.  
OC Viruses; dsDNA viruses, no RNA stage; Caudovirales; Podoviridae;  
OC P22-like viruses.  
OX NCBI\_TaxID=10754;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=85155495; PubMed=6241581;  
RA Backhaus H., Petri J.B.;  
RT "Sequence analysis of a region from the early right operon in phage  
RT P22 including the replication genes 18 and 12.";  
RL Gene 32:289-303 (1984).  
RN [2]  
RP SEQUENCE FROM N.A.  
RA Kropinski A.M.B., Vanderbyl C.S.;  
RT "The completed sequence of genome of Salmomella phage P22";  
RL Submitted (DEC-1999) to the EMBL/Genbank/DBJ databases.  
CC -1- SIMILARITY: THIS PROTEIN IS RELATED TO THE REPLICATION PROTEIN O  
CC OF BACTERIOPHAGE LAMBDA.  
-----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
-----  
DR EMBL; M10074; AAA32275.1; -.  
DR EMBL; AF217253; AAF75028.1; -.  
DR PIR; A04293; Z8BPC2.  
KW DNA replication; Early protein.  
SQ SEQUENCE 271 AA; 30557 MW; 084F0E8918FDACA2 CRC64;  
  
Query Match 31.2%; Score 5; DB 1; Length 271;

Best Local Similarity 100.0%; Pred. No. 1.1e+02;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
QY 7 DNDRD 11  
Db 245 DNDRD 249  
  
RESULT 31  
YPV2\_METTF STANDARD; PRT; 284 AA.  
ID \_YPV2\_METTF  
AC P29571;  
DT 01-APR-1993 (Rel. 25, Created)  
DT 01-APR-1993 (Rel. 25, Last sequence update)  
DT 01-APR-1993 (Rel. 25, Last annotation update)  
DE Hypothetical 33.2 kDa protein (ORF2).  
OS Methanobacterium thermoformicicum.  
OC Plasmid pFV1.  
OC Archaea; Euryarchaeota; Methanobacteria; Methanobacteriales;  
OC Methanobacteriaceae; Methanothermobacter.  
OX NCBI\_TaxID=145262;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=DSM 3848 / THF;  
RX MEDLINE=93126090; PubMed=1336177;  
RA Noelling J., van Eeden F.J.M., Eggen R.I.T., de Vos W.M.;  
RT "Modular organization of related Archaeal plasmids encoding different  
RT restriction-modification systems in Methanobacterium  
RT thermoformicicum.";  
RL Nucleic Acids Res. 20:6501-6507 (1992).  
CC -1- SIMILARITY: TO THE CORRESPONDING ORF IN PZL1.  
-----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
-----  
DR EMBL; X68366; CAA48426.1; -.  
DR PIR; S26437; S26437.  
DR PIR; S30302; S30302.  
KW Hypothetical protein; Plasmid.  
SQ SEQUENCE 284 AA; 33212 MW; 0E48AC802417FE CRC64;  
  
Query March 31.2%; Score 5; DB 1; Length 284;  
Best Local Similarity 100.0%; Pred. No. 1.1e+02;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
QY 4 GRDDN 8  
Db 108 GRDDN 112  
  
RESULT 32  
YPV2\_METTF STANDARD; PRT; 284 AA.  
ID \_YPV2\_METTF  
AC P29572;  
DT 01-APR-1993 (Rel. 25, Created)  
DT 01-APR-1993 (Rel. 25, Last sequence update)  
DT 01-APR-1993 (Rel. 25, Last annotation update)  
DE Hypothetical 33.1 kDa protein (ORF2').  
OS Methanobacterium thermoformicicum.  
OC Plasmid pFV1.  
OC Archaea; Euryarchaeota; Methanobacteria; Methanobacteriales;  
OC Methanobacteriaceae; Methanothermobacter.  
OX NCBI\_TaxID=145262;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=DSM 3720 / Z-245;  
RX MEDLINE=93126090; PubMed=1336177;  
RA Noelling J., van Eeden F.J.M., Eggen R.I.T., de Vos W.M.;

```
RT "Modular organization of related Archaeal plasmids encoding different
RT restriction-modification systems in Methanobacterium
RT thermophilicum."
RL Nucleic Acids Res. 20:6501-6507(1992).
CC -1- SIMILARITY: TO THE CORRESPONDING ORF IN PFV1.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by, and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
DR EMBL; X68367; CAA48441.1; -.
DR PIR; S26451; S26451.
DR PIR; S30317; S30317.
KW Hypothetical protein; Plasmid.
SQ SEQUENCE 284 AA; 33107 MW; EBAC63243FDBE897 CRC64;

Query Match 31.2%; Score 5; DB 1; Length 284;
Best Local Similarity 100.0%; Pred. No. 1.1e+02;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4 GRDDN 8
    |||||
Db 108 GRDDN 112

RESULT 33
MYO2 HUMAN STANDARD; PRT; 288 AA.
ID MYO2_HUMAN
AC 014732;
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Myo-inositol 1(or 4)-monophosphatase 2 (EC 3.1.3.25) (IMPAse 2) (IMP
DE 2) (Inositol monophosphatase 2) (Myo-inositol monophosphatase A2).
GN IMPA2.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxID=9606;
[1]
RP SEQUENCE FROM N.A.
RX MEDLINE=97463449; PubMed=9322233;
RA Yoshikawa T.; Turner G.; Saterling L.E.; Sanders A.R.;
RA Devera-Wadleigh S.D.;
RT "A novel human myo-inositol monophosphatase gene, IMP.18p, maps to a
RT susceptibility region for bipolar disorder.";
RL Mol. Psych. 2:393-397(1997).
RN [2]
RN SEQUENCE FROM N.A.
RX MEDLINE=20284188; PubMed=10822345;
RA Stoholt G.; Gulbrandsen A.K.; Lovlie R.; Berle J.; Molven A.;
RA Stoen V.M.;
RT "A human myo-inositol monophosphatase gene (IMPA2) localized in a
RT putative susceptibility region for bipolar disorder on chromosome
RT 18p11.2: genomic structure and polymorphism screening in manic-
RT depressive patients.";
RL Mol. Psych. 5:172-180(2000).
CC -1- CATALYTIC ACTIVITY: Myo-inositol 1-monophosphate + H(2)O = myo-
CC inositol + phosphate.
CC -1- PATHWAY: KEY ENZYME OF THE PHOSPHATIDYL INOSITOL SIGNALING
CC PATHWAY.
CC -1- SIMILARITY: BELONGS TO THE INOSITOL MONOPHOSPHATASE FAMILY.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by, and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
```

```
CC or send an email to license@isb-sib.ch).
-----
DR EMBL; AF014398; AAB70915.1; -.
DR EMBL; AF157102; AAD40683.1; -.
DR EMBL; AF157096; AAD40683.1; JOINED.
DR EMBL; AF157097; AAD40683.1; JOINED.
DR EMBL; AF157098; AAD40683.1; JOINED.
DR EMBL; AF157099; AAD40683.1; JOINED.
DR EMBL; AF157100; AAD40683.1; JOINED.
DR EMBL; AF157101; AAD40683.1; JOINED.
DR HSSP; P29218; IMF.
DR Genew; HGNC:6051; IMPA2.
DR MIM; 605922; -.
DR InterPro; IPR000760; Inositol_P.
DR Pfam; PF00459; Inositol_P; 1.
DR PROSITE; PS00629; IMP_1; 1.
DR PROSITE; PS00630; IMP_2; 1.
KW Hydrolyase.
SQ SEQUENCE 288 AA; 31321 MW; 30D83FA339AC69B4 CRC64;

Query Match 31.2%; Score 5; DB 1; Length 288;
Best Local Similarity 100.0%; Pred. No. 1.1e+02;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 YGRDD 7
    |||||
Db 282 YGRDD 286

RESULT 34
YNP9 YEAST STANDARD; PRT; 289 AA.
ID YNP9_YEAST
AC P53895;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 01-OCT-1996 (Rel. 34, Last annotation update)
DE Hypothetical 33.0 kDa protein in YGP1-YCK2 intergenic region.
DE YNL159C OR N1735.
GN Saccharomyces cerevisiae (Baker's yeast).
OS Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomyces.
OX NCBI_TaxID=4932;
[1]
RP SEQUENCE FROM N.A.
RX STRAIN=S288c / FY1679;
RX MEDLINE=96287653; PubMed=8686380;
RA Nasr F.; Becam A.-M.; Herbert C.J.;
RT "The sequence of 36.8 kb from the left arm of chromosome XIV reveals
RT 24 complete open reading frames: 18 correspond to new genes, one of
RT which encodes a protein similar to the human myotonic dystrophy
RT kinase.";
RL Yeast 12:169-175(1996).
RN [2]
RN SEQUENCE FROM N.A.
RX MEDLINE=271435; CAA96046.1; -.
DR EMBL; Z71435; CAA96046.1; -.
DR SGD; S0005103; YNL159C.
KW Hypothetical protein; Transmembrane.
KW TRANSMEM 122 142 POTENTIAL.
FT TRANSMEM 248 268 POTENTIAL.
SQ SEQUENCE 289 AA; 32985 MW; 95C425488F1E3724 CRC64;

Query Match 31.2%; Score 5; DB 1; Length 289;
Best Local Similarity 100.0%; Pred. No. 1.1e+02;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```

OY      5 RDDND 9
        |||||
        15 RDDND 19

Db      15 RDDND 19

RESULT 35
Y42A RHISN
ID      Y42A RHISN      STANDARD;      PRT;      295 AA.
AC      P55728;
DT      01-NOV-1997 (Rel. 35, Created)
DT      01-NOV-1997 (Rel. 35, Last sequence update)
DT      01-NOV-1997 (Rel. 35, Last annotation update)
DE      Hypothetical 32.9 Kda protein Y42A.
GN      Y42A.
OS      Rhizobium sp. (strain NGR234).
OC      Plasmid bym DNGR234a.
OC      Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
OC      Rhizobiaceae; Rhizobium.
OX      NCBI_TaxID=394;
RN      [1]
RP      SEQUENCE FROM N.A.
RX      MEDLINE=97305956; PubMed=9163424;
RA      Freiberg C.A., Pelly R., Bairoch A., Broughton W.J., Rosenthal A.,
RT      Petrec X.;
RL      "Molecular basis of symbiosis between Rhizobium and legumes.";
CC      Nature 387:394-401(1997).
CC      -1 SIMILARITY: NONE OBVIOUS OUTSIDE OF SIMILARITY WITH Y4RL AND Y4RM
CC      AS WELL AS Y4IE.
-----
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC      or send an email to license@isb-sib.ch).
-----
CC      -----
DR      EMBL; AE00108; AAB91959.1; -.
KW      Hypothetical protein; Plasmid.
FT      SIMILAR 34 115 HIGHLY SIMILAR TO Y4RM 258-339.
FT      SIMILAR 133 231 HIGHLY SIMILAR TO Y4RL 1-99.
SQ      SEQUENCE 295 AA; 32927 MW; C92240D19F38782 CRC64;

Query Match      31.2%; Score 5; DB 1; Length 295;
Best Local Similarity 100.0%; Pred. No. 1.2e+02;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY      1 VGYGR 5
        |||||
        58 VGYGR 62

Db      58 VGYGR 62

RESULT 36
Y103 HUMAN
ID      Y103 HUMAN      STANDARD;      PRT;      297 AA.
AC      Q15006;
DT      15-JUL-1998 (Rel. 36, Created)
DT      15-JUL-1998 (Rel. 36, Last sequence update)
DT      15-JUN-2002 (Rel. 41, Last annotation update)
DE      Hypothetical protein KIAA0103.
GN      KIAA0103.
OS      Homo sapiens (Human).
OC      Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC      Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX      NCBI_TaxID=9606;
RN      [1]
RP      SEQUENCE FROM N.A.
RC      TISSUE=Bone marrow;
RX      MEDLINE=95308325; PubMed=7788527;
RA      Nagase T., Miyajima N., Tanaka A., Sazuka T., Seki N., Sato S.,
RA      Tabata S., Ishikawa K.-I., Kawarayashi Y., Kotani H., Nomura N.;
RT      "Prediction of the coding sequences of unidentified human genes. III.
RT      The coding sequences of 40 new genes (KIAA0081-KIAA0120) deduced by

```

```

RT      analysis of cDNA clones from human cell line KG-1.";
RL      DNA Res. 2:37-43(1995).
RN      [2]
RP      SEQUENCE FROM N.A.
RC      TISSUE=Urinary bladder;
RA      Strauberg R.;
RL      Submitted (JUN-2002) to the EMBL/GenBank/DBJ databases.
CC      -----
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC      or send an email to license@isb-sib.ch).
-----
CC      -----
DR      EMBL; D14659; BAA03493.1; -.
DR      EMBL; BC021667; AAH21667.1; -.
DR      InterPro; IPR001440; IPR.
DR      Pfam; PF00515; TPR; 2.
KW      Hypothetical protein.
SQ      SEQUENCE 297 AA; 34833 MW; CA5903ARC3B3E36A CRC64;

Query Match      31.2%; Score 5; DB 1; Length 297;
Best Local Similarity 100.0%; Pred. No. 1.2e+02;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY      3 YGRDD 7
        |||||
        66 YGRDD 70

Db      66 YGRDD 70

RESULT 37
CCDC CAEEL
ID      CCDC CAEEL      STANDARD;      PRT;      302 AA.
AC      P17657;
DT      01-AUG-1990 (Rel. 15, Created)
DT      01-NOV-1990 (Rel. 16, Last sequence update)
DT      15-JUN-2002 (Rel. 41, Last annotation update)
DE      Cuticle collagen dpy-13.
GN      Dpy-13 OR Dpy-16 OR F30B5.1.
OS      Caenorhabditis elegans.
OC      Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC      Rhabditidae; Pelodierinae; Caenorhabditis.
OX      NCBI_TaxID=6239;
RN      [1]
RP      SEQUENCE FROM N.A.
RX      MEDLINE=89028668; PubMed=2846184;
RA      von Mende N., Bird D.M., Albert P.S., Riddle D.L.;
RT      "dpy-13: a nematode collagen gene that affects body shape.";
RL      Cell 55:567-576(1988).
RN      [2]
RP      SEQUENCE FROM N.A.
RC      STRAIN=Bristol N2;
RA      Waterston R.;
RL      Submitted (DEC-1995) to the EMBL/GenBank/DBJ databases.
RN      [3]
RX      VARIANT Dpy13 (E225).
RX      MEDLINE=93013043; PubMed=1398138;
RA      Bird D.M.;
RT      "Sequence comparison of the Caenorhabditis elegans dpy-13 and col-34
RT      genes, and their deduced collagen products.";
RL      Gene 120:261-266(1992).
CC      -1- FUNCTION: NEMATODE CUTICLES ARE COMPOSED LARGELY OF COLLAGEN-LIKE
CC      PROTEINS. THE CUTICLE FUNCTIONS BOTH AS AN EXOSKELETON AND AS A
CC      BARRIER TO PROTECT THE WORM FROM ITS ENVIRONMENT.
CC      -1- SUBUNIT: COLLAGEN POLYPEPTIDE CHAINS ARE COMPLEXED WITHIN THE
CC      CUTICLE BY DISULFIDE BONDS AND OTHER TYPES OF COVALENT CROSS-
CC      LINKS.
CC      -1- DISEASE: MUTATIONS IN Dpy-13 AFFECTS THE BODY SHAPE.
CC      -1- SIMILARITY: BELONGS TO THE CUTICULAR COLLAGEN FAMILY.
-----
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration

```

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC -----  
 DR EMBL; M23559; AAA27994.1; -.  
 DR EMBL; U42437; AAA83499.1; -.  
 DR PIR; A31921; A31921.  
 DR Wormpep; F3085.1; CE04462.  
 DR InterPro; IPR002486; Col\_cuticle\_N.  
 DR InterPro; IPR000087; Collagen.  
 DR Pfam; PF01391; Collagen; 2.  
 DR Pfam; PF01484; Col\_cuticle\_N; 1.  
 KW Cuticle; Connective tissue; Repeat; Multigene family; Collagen.  
 FT DOMAIN 106 135 TRIPLE-HELICAL REGION.  
 FT DOMAIN 154 210 TRIPLE-HELICAL REGION.  
 FT DOMAIN 219 278 TRIPLE-HELICAL REGION.  
 FT VARIANT 205 205 G -> E (IN DPLY3(B225)).  
 SQ SEQUENCE 302 AA; 30052 MW; 16B5CDAP14E0F0E5 CRC64;

Query Match 31.2%; Score 5; DB 1; Length 302;  
 Best Local Similarity 100.0%; Pred. No. 1.2e+02;  
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 3 YGRDD 7  
 Db 82 YGRDD 86

RESULT 38  
 SITE HUMAN STANDARD; PRT; 336 AA.  
 AC Q9BVT7;  
 DT 15-JUN-2002 (Rel. 41, Created)  
 DT 15-JUN-2002 (Rel. 41, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE Alpha-N-acetylglucosaminide alpha-2,6-sialyltransferase V (GDI alpha  
 DE synthase) (GALNAC alpha-2,6-sialyltransferase V) (ST6GALNAC V)  
 DE (Stalyltransferase 7E).  
 GN SIAT7E.  
 OS Homo sapiens (Human).  
 CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 CC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
 CX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Ieribashi T., Kanehori K., Yosida M., Matanabe S., Ishida S., Ono Y.,  
 RA Hota T., Hiraoka S., Murakawa K., Takiguchi S., Kusano J.,  
 RA Matanabe M., Fujimori K., Tanai H., Ishida M., Yamashita H., Chiba Y.,  
 RA Sugiyama T., Irie R., Otsuki T., Sato H., Wakamatsu A., Ishii S.,  
 RA Yamamoto J., Isono Y., Kawai-Hio Y., Saito K., Nishikawa T.,  
 RA Kimura K., Matsuo K., Nakamura Y., Sekito M., Kikuchi H., Kanda K.,  
 RA Wagaatsuma M., Takahashi-Fujii A., Oshima A., Sugiyama A., Kawakami B.,  
 RA Suzuki Y., Sugano S., Nagahara K., Masuno Y., Nagai K., Isogai T.,  
 RT "NBD human cDNA sequencing project."  
 RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Eye;  
 RA Straubeberg R.;  
 RL Submitted (DEC-2000) to the EMBL/GenBank/DBJ databases.  
 CC -1- FUNCTION: Involved in the biosynthesis of ganglioside Gd1a from  
 CC glycoproteins (By similarity).  
 CC -1- PATHWAY: Glycosylation.  
 CC -1- SUBCELLULAR LOCATION: Type II membrane protein. Golgi (By  
 CC similarity).  
 CC -1- SIMILARITY: BELONGS TO THE GLYCOSYLTRANSFERASE FAMILY 29.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC -----  
 DR EMBL; AK056241; BAB7127.1; -.  
 DR EMBL; BC001201; AAH01201.1; -.  
 DR InterPro; IPR001675; GT\_29.  
 DR Pfam; PF00777; Glyco transf 29; 1.  
 KW Transferase; Glycosyltransferase; Glycoprotein; Transmembrane;  
 KW Signal-anchor; Golgi stack.  
 FT DOMAIN 1 8  
 FT TRANSMEM 9 29  
 FT SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN)  
 FT LUMENAL, CATALYTIC (POTENTIAL).  
 FT DOMAIN 30 336  
 FT CARBOHYD 38 49 POLY-GLN.  
 FT CARBOHYD 137 137 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 161 161 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 SQ SEQUENCE 336 AA; 38443 MW; F5FDD43D45CA11CE CRC64;

Query Match 31.2%; Score 5; DB 1; Length 336;  
 Best Local Similarity 100.0%; Pred. No. 1.3e+02;  
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 2 GYGRD 6  
 Db 130 GYGRD 134

RESULT 39  
 MRAM MYCLE STANDARD; PRT; 340 AA.  
 AC O69560;  
 DT 30-MAY-2000 (Rel. 39, Created)  
 DT 30-MAY-2000 (Rel. 39, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE S-adenosyl-methyltransferase mrav (EC 2.1.1.-).  
 GN MRAM OR M10906 OR M1CB268.10C.  
 OS Mycobacterium leprae.  
 CC Bacteria; Actinobacteria; Actinobacteria (class); Actinobacteridae;  
 CC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.  
 CX NCBI\_TaxID=1769;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=TN;  
 RX MEDLINE=21128732; PubMed=11234002;  
 RA Cole S.T., Eiglmeyer K., Parkhill J., James K.D., Thomson N.R.,  
 RA Wheeler P.R., Honore N., Garnier T., Churcher C., Harris D.,  
 RA Mungall K., Basham D., Brown D., Chillingworth T., Connor R.,  
 RA Davies R.M., Devlin K., Duthoy S., Felwell T., Fraser A., Hamlin N.,  
 RA Holtroyd S., Hornsby T., Jagers K., Lacroix C., Maclean J., Moule S.,  
 RA Murphy L., Oliver K., Quail M.A., Rajandream M.A., Rutherford K.M.,  
 RA Rutter S., Seeger K., Simon S., Simmonds M., Skelton J., Squares R.,  
 RA Squares S., Stevens K., Taylor K., Whitehead S., Woodward J.R.,  
 RA Barrell B.G.;  
 RT "Massive gene decay in the leprosy bacillus."  
 RL Nature 409:1007-1011 (2001).  
 CC -1- FUNCTION: Exhibits a S-adenosyl-dependent methyltransferase  
 CC activity (By similarity).  
 CC -1- SIMILARITY: BELONGS TO THE MRAM FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC -----  
 DR EMBL; AL022602; CAA18676.1; -.  
 DR EMBL; AL583920; CAC31287.1; ALT\_INIT.  
 DR Leproma; M10906; -.  
 CC -----

DR InterPro; IPR002903; Bac\_Mettrifree.  
DR InterPro; IPR000051; SAM\_Bind.  
DR Pfam; PF01795; Methyltransf\_5; 1.  
DR ProDom; PD004685; Bac\_Mettrifree; 1.  
DR TIGRFAMs; TIGR00006; UPF0117; 1.  
KM Transferrase; Methyltransferase; Complete proteome.  
SQ SEQUENCE 340 AA; 37146 MW; 2A64E02E040FD276 CRC64;

Query Match 31.2%; Score 5; DB 1; Length 340;  
Best Local Similarity 100.0%; Pred. No. 1.3e+02;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 9 DRDPs 13  
Db 65 DRDPs 69

RESULT 40  
SAPD\_HAEIN STANDARD; PRT; 349 AA.  
ID SAPD\_HAEIN  
AC P45288;  
DT 01-NOV-1995 (Rel. 32, Created)  
DT 01-NOV-1995 (Rel. 32, Last sequence update)  
DT 15-JUN-2002 (Rel. 41, Last annotation update)  
DE Peptide transport system ATP-binding protein sapd.  
OS Haemophilus influenzae.  
OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;  
OC Haemophilus.  
OX NCBI\_TaxID=727;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=Rd / KW20 / ATCC 51907;  
RX MEDLINE=95350630; PubMed=7542800;  
RA Fleischmann R.D., Adams M.D., White O., Clayton R.A., Kirkness E.F.,  
RA Keriavage A.R., Bult C.J., Tomb J.-F., Dougherty B.A., Merrick J.M.,  
RA McKenney K., Sutton G., Fitzhugh W., Fields C.A., Gocayne J.D.,  
RA Scott J.D., Shirley R., Liu L.-I., Glodek A., Kelley J.M.,  
RA Weidman J.F., Phillips C.A., Spriggs T., Hedblom E., Cotton M.D.,  
RA Uterback T.R., Hanna M.C., Nguyen D.T., Sauder D.M., Brandon R.C.,  
RA Fine L.D., Fritchman J.L., Fuhrmann J.L., Geoghagen N.S.M.,  
RA Gnehm C.L., McDonald L.A., Small K.V., Fraser C.M., Smith H.O.,  
RA Venter J.C.;  
RT "Whole-genome random sequencing and assembly of Haemophilus influenzae  
Rd.";  
RL Science 269:496-512(1995).  
CC -1- FUNCTION: INVOLVED IN A PEPTIDE INTAKE TRANSPORT SYSTEM THAT  
CC PLAYS A ROLE IN THE RESISTANCE TO ANTIMICROBIAL PEPTIDES  
CC (BY SIMILARITY).  
CC -1- SUBCELLULAR LOCATION: Inner membrane-associated (Potential).  
CC -1- SIMILARITY: BELONGS TO THE ABC TRANSPORTER FAMILY.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
CC EMBL; U32837; AAC3288.1; -.  
DR TIGR; H11641; -.  
DR InterPro; IPR003439; ABC\_transportr.  
DR Pfam; PF00005; ABC\_tran; 1.  
DR ProDom; PD000006; ABC\_transportr; 1.  
DR PROSITE; PS00211; ABC\_TRANSPORTER; FALSE NEG.  
KW Peptide transport; transport; inner membrane; ATP-binding;  
KW Complete proteome.  
FT NP BIND 40 47 ATP (POTENTIAL).  
SQ SEQUENCE 349 AA; 39478 MW; 25FD3241F1570D7A CRC64;

Query Match 31.2%; Score 5; DB 1; Length 349;  
Best Local Similarity 100.0%; Pred. No. 1.4e+02;

Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
OY 11 DPSRK 15  
Db 103 DPSRK 107

Search completed: April 9, 2003, 13:13:10  
Job time : 15.8033 secs

GenCore version 5.1.4 p5 4578  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 8, 2003, 06:19:34 ; Search time 3360 Seconds  
(without alignments)  
12793.108 Million cell updates/sec

Title: US-10-020-441-1  
Perfect score: 1477  
Sequence: 1 gtgtcaacccgtgtgacacg.....ttgtacgtcccaatttga 1477

Scoring table: IDENTITY NUC  
Gapop 10.0, Gapext 1.0

Searched: 2054640 seqs, 14551402878 residues  
Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

GenEmbl:  
1: gb\_ba:\*  
2: gb\_hg:\*  
3: gb\_in:\*  
4: gb\_cm:\*  
5: gb\_ov:\*  
6: gb\_pat:\*  
7: gb\_ph:\*  
8: gb\_pl:\*  
9: gb\_pr:\*  
10: gb\_ro:\*  
11: gb\_sy:\*  
12: gb\_un:\*  
13: gb\_vl:\*  
14: gb\_da:\*  
15: em\_fun:\*  
16: em\_hum:\*  
17: em\_in:\*  
18: em\_mu:\*  
19: em\_om:\*  
20: em\_or:\*  
21: em\_ov:\*  
22: em\_pat:\*  
23: em\_ph:\*  
24: em\_pl:\*  
25: em\_ro:\*  
26: em\_ets:\*  
27: em\_un:\*  
28: em\_vl:\*  
29: em\_hg\_hum:\*  
30: em\_hg\_inv:\*  
31: em\_hg\_other:\*  
32: em\_hg\_mus:\*  
33: em\_hg\_pln:\*  
34: em\_hg\_xod:\*  
35: em\_hg\_mam:\*  
36: em\_hg\_vrt:\*  
37: em\_hg\_hum:\*  
38: em\_hg\_mus:\*  
39: em\_hg\_hum:\*  
40: em\_hg\_mus:\*  
41: em\_hg\_other:\*

score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1475	99.9	1477	3	SMHPIELAS
2	1198.2	81.1	3045	3	SMU31768
3	450.6	30.5	1163	3	SMU31769
4	332.4	22.5	1152	3	SCMCROT
5	287.2	19.4	904	3	AF510334
6	279.2	18.9	943	3	AF510341
7	248.8	16.8	419	3	AF510342
8	240.8	16.3	431	3	AF510343
9	183.4	9.7	886	3	AF510340
10	88.2	6.0	441	3	AF510344
11	65.8	4.5	1630	2	AC007547
12	63.4	4.3	201861	2	AC103344
13	61.4	4.2	199649	10	AL669825
14	61.2	4.1	236562	2	AL772338
15	60.6	4.1	145534	2	AC115200
16	58.8	4.0	248706	2	AC095279
17	57.8	3.9	105899	2	AC098634
18	57.8	3.9	179590	2	AC130952
19	57.2	3.9	107628	2	AC131220
20	57.2	3.9	132168	2	AC119336
21	57.2	3.9	193437	2	AC126667
22	57.2	3.9	172387	2	AC128718
23	56.4	3.8	159876	2	AC113711
24	56.2	3.8	73678	2	AC123564
25	56	3.8	168382	2	AC098913
26	56	3.8	223854	10	AL606962
27	55.8	3.8	171870	2	AC111443
28	55.6	3.8	211969	10	AL671970
29	55.4	3.8	155682	2	AC131189
30	55.4	3.8	174697	2	AC128816
31	55.4	3.8	185413	2	AC119009
32	55.4	3.8	188030	2	AC114869
33	55.2	3.7	167892	2	AC110512
34	55.2	3.7	176192	2	AC109062
35	55.2	3.7	215212	2	AC125154
36	54.8	3.7	15446	9	AY071830
37	54.8	3.7	65472	2	AC120787
38	54.8	3.7	156238	2	AC104862
39	54.8	3.7	158661	9	HS271736
40	54.8	3.7	176469	2	AC121034
41	54.8	3.7	183559	2	AC129788
42	54.4	3.7	164582	2	CNS08C9P
43	54.4	3.7	170920	2	AC120911
44	54.4	3.7	176868	2	AC097668
45	54.2	3.7	114260	2	AC098956

ALIGNMENTS

RESULT 1	LOCUS	SMHPIELAS	1477 bp	DNA	linear	INV 29-MAR-1996
DEFINITION	S. mansoni	elastase HPI gene.				
ACCESSION	Z70296					
VERSION	Z70296.1	GI:1240057				
KEYWORDS	elastase.					
SOURCE	Schistosoma mansoni.					
ORGANISM	Schistosoma mansoni					
REFERENCE	Strigoida; Metazoa; Platyhelminthes; Trematoda; Digenea;					
AUTHORS	Sayers, J.R., Price, H.P. and Doenhoff, M.J.					
JOURNAL	Unpublished					
REFERENCE	2 (bases 1 to 1477)					





ORGANISM Schistosoma mansoni  
Eukaryota; Metazoa; Platyhelminthes; Trematoda; Digenea;  
Sturigidida; Schistosomatidae; Schistosomatidae; Schistosoma.  
REFERENCE 1 (bases 1 to 3045)  
AUTHORS Pierrot, C., Capron, A. and Khalife, J.  
TITLE Cloning and characterization of two genes encoding Schistosoma  
mansoni elastase  
JOURNAL Mol. Biochem. Parasitol. 75 (1), 113-117 (1995)  
MEDLINE 96362066  
PUBMED 8720180  
REFERENCE 2 (bases 1 to 3045)  
AUTHORS Pierrot, C.  
TITLE Direct Submission  
JOURNAL Submitted (18-JUL-1995) Christine Pierrot, CIBP, INSERM U 167 -  
Institut Pasteur de Lille, 1, rue du Professeur Calmette, L111e  
59019, France  
FEATURES  
Location/Qualifiers  
1..3045  
/organism="Schistosoma mansoni"  
/strain="Puerto rico"  
/db\_xref="taxon:6183"  
join(1..1465,1588..1935,2525..>2910)  
/product="elastase"  
1..1231  
5' UTR 1153..1160  
CAAT\_signal 1162..1167  
TATA\_signal join(1232..1465,1588..1935,2525..2767)  
CDS /codon\_start=1  
/product="elastase"  
/protein\_id="AAC46967.1"  
/translation="MSNRMRPLVVTLLFTYCLTFRVSTWLRSGEPVQHRTEPFIAR  
LTERMTGCTSLVTRAVLTAGHCVCSPYPIRVLCPLQVDFLRNDQGHOPHS  
GVKRAPGMSCSMARBRPPIAQTLSGSDIAIWLQWVNIQSGITVSLDQADIP  
PGRVFTVYGRDNDPRDPRKRGKILKKEGLVGRATIMECRHATNPNICVTAQON  
FGOLPAPGDSGPIPLSPQGPVLGVSHGVTLPMLDILVYASVARMIDVRNIN"  
3' UTR 2768..>2910  
polya\_signal 2886..2891  
polya\_signal 2906..2910  
BASE COUNT 851 a 579 c 691 g 924 t  
ORIGIN  
Query Match 81.1%; Score 1198.2; DB 3; Length 3045;  
Best Local Similarity 94.1%; Pred. No. 0;  
Matches 1401; Conservative 1; Mismatches 59; Indels 28; Gaps 14;  
QY 1 GTGTCAACCTGTTGATACGTAGTGTGAACCTGTGCACACCGCACTGAATCCCGTTC 60  
DB 1295 GTGTCAACCTGTTGATACGTAGTGTGAACCTGTGCACACCGCACTGAATCCCGTTC 1354  
QY 61 ATGCATCTTAAGCAGAGAGAACAAATGTGTACAGSTTCACTAGTCTCAAGAGAGCA 120  
DB 1355 ATGCATCTTAAGCAGAGAGAACAAATGTGTACAGSTTCACTAGTCTCAAGAGAGCA 1414  
QY 121 GTACTCAGAGTGTGTCATTTGTTGTCTCACCATTGCCAGTATCGGGTAAGATCG- 179  
DB 1415 GTACTCAGAGTGTGTCATTTGTTGTCTCACCATTGCCAGTATCGGGTAAGATCGC 1474  
QY 180 ACTGAAACACATTTGTTGTCATGTAATGATTTTCAACAGACAGTGTTCGTGTCC 239  
DB 1475 ACTGAAACACATTTGTTGTCATGTAATGATTTTCAACAGACAGTGTTCGTGTCC 1534  
QY 240 GTGCGTTGTGTTGTATGTCAGTGGGTTTGATAGATTTTAATGTCATATACCTGT 299  
DB 1353 GTGCGTTGTGTTGTATGTCAGTGGGTTTGATAGATTTTAATGTCATATACCTGT 1593  
QY 300 CTATATTCAGGTTTCATTTCTCACACTGAGAAATGSCGACCAACAGGATCATCACCA 359  
DB 1594 C-TATATTCAGGTTTCATTTCTCACACTGAGAAATGSCGACCAACAGGATCATCACCA 1652  
QY 360 ACCGCTCTGAAGTTAAGTGCGACACAGATACATGCCCTTGTATGTGCGACGACAG 419  
DB 1653 ACCATCTGAGATTAAAGTGCGACACAGATATATGCCCTTGTATGTGCGACGAGAG 1712

QY 420 GAGACCAATCGCAGACAGACACTCAGTGAATTCGATTTTCAATTTGATGCTGGCTCAAT 479  
DB 1713 GAGACCAATCGCAGACCAACACTCAGTGAATTCGACACTTCAATTTGATGCTGGCTCAAT 1772  
QY 480 GGTCACTTACAGAGTGAATCAGATGATCATGTCGCCACACCATGTGATATCCGCC 539  
DB 1773 GGTCACTTACAGAGTGAATCAGATGATCATGTCGCCACACCATGTGATATCCGCC 1832  
QY 540 ACCGTGAATCGGTGTTTCATTTGTTGTTATGGAAGGATGATACGACCGGATCCGCT 599  
DB 1833 ACCGTGAATCGGTGTTTCATTTGTTGTTATGGAAGGATGATACGACCGGATCCGCT 1892  
QY 600 ACGTAAGATGCGTGAATATTTGAAGAAA-GTGAGTTGTTGTTGAATTAACGACATGCTC 658  
DB 1893 TCGTAATAATGCTGAATTTTGAAGAAAAGTGAGTTGATGTTGTCACAAACATGATCTC 1952  
QY 659 AGTC-----GTCACTCAGATGTCAGTCACTTATCTGTGTGTCGTGATCTGTTGTGT 713  
DB 1953 AGTCAGTCAGTTTGTCACTAGTCACTGATCTGATCTGTGTGTCGTGATCTGTTGTGT 2012  
QY 714 GTCTGTCTGTACCTGATCCGTTGTTGTTATGTTGTCAGACCTTGAATTAACAACGT 773  
DB 2013 GTCTGTCTGTACCTGATCCGTTGTTGTTATGTTGTCAGACCTTGAATTAACAACGT 2072  
QY 774 GTTTGATGACTTTGTGACAGTTCAGTACAGAGTATTTCCATCCGTCATTTGTGTG 833  
DB 2073 GTTTGATGAGC-TTGTGACAGTTCAGTACAGAGTATTTCCATCCGTCATTTGTGTG 2130  
QY 834 GTGAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGT 892  
DB 2131 GTGAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGT 2190  
QY 893 ATGTGAATGGAATGATTTGACACCTTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 952  
DB 2191 ATGTGAATGGAATGATTTGACACCTTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2250  
QY 953 CGATGGAAGTGTGTTG-TGTACATGAAGTGAAGGCGTCAATGTTGTTGATGTTGTTG 1011  
DB 2251 CGATGGAAGTGTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTG 2310  
QY 1012 AGAGTGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGT 1069  
DB 2311 AGAGTGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGT 2370  
QY 1070 TGCACATA-TTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1128  
DB 2371 TGCACATA-TTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 2430  
QY 1129 AATTCATTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTT 1188  
DB 2431 AATTCATTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTT 1248  
QY 1189 CGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTT 1248  
DB 2486 CTGACACTCTACGTGATCTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTT 2538  
QY 1249 AATGGAATGCCGACATGCGACCAATGCAATCTATATGTGTGAAGAGAGAGTCAATTT 1308  
DB 2539 AATGGAATGCCGACATGCGACCAATGCAATCTATATGTGTGAAGAGAGAGTCAATTT 2598  
QY 1309 CGGACAGTTACCGGCTCCAGGTTGACAGTGTGAGCCTCTCCCATCCCTTCAAGAGTCC 1368  
DB 2599 CGGACAGTTACCGGCTCCAGGTTGACAGTGTGAGCCTCTCCCATCCCTTCAAGAGTCC 2658  
QY 1369 AGTACTCGGTGTGTATCATATGTTGTCACTGCGAAAACCTTCCGATATCATTTGGA 1428  
DB 2659 AGTACTCGGTGTGTATCATATGTTGTCACTGCGAAAACCTTCCGATATCATTTGGA 2718  
QY 1429 GTATGCCAGTGTGCTGAAGATTTGATTTTGTACGCTCCAAATTTTGA 1477  
DB 2719 GTATGCCAGTGTGCTGAAGATTTGATTTTGTACGCTCCAAATTTTGA 2767

```

RESULT 3
LOCUS SMU31769 1163 bp DNA linear INV 28-JAN-1996
DEFINITION Schistosoma mansoni elastase gene, 1163 bp clone, complete cds.
ACCESSION U31769
VERSION U31769.1 GI:1103830
KEYWORDS
SOURCE blood fluke.
ORGANISM Schistosoma mansoni
Eukaryota; Metazoa; Platyhelminthes; Trematoda; Digenea;
Strigoidida; Schistosomatidae; Schistosomatidae; Schistosoma.
REFERENCE
1 (bases 1 to 1163)
Pierrot,C., Capron,A. and Khalife,J.
Cloning and characterization of two genes encoding Schistosoma
mansoni elastase
Mol. Biochem. Parasitol. 75 (1), 113-117 (1995)
JOURNAL MEDLINE 96362066
PUBMED 8720180
REFERENCE 2 (bases 1 to 1163)
Pierrot,C.
Direct Submision
AUTHORS Submitted (18-JUN-1995) Christine Pierrot, CIBP, INSERM U167 -
JOURNAL Institut Pasteur de Lille, 1, rue du Professeur Calmette, Lille,
59019 France
FEATURES
source
1..1163
/organism="Schistosoma mansoni"
/strain="Puerto Rico"
/db_xref="taxon:6183"
join(<1..234,360..668,933..>1163)
join(1..234,360..668,933..1163)
/codon_start=1
/product="elastase"
/protein_id="AAC46968.1"
/db_xref="GI:1103831"
/translation="MSNMRFLVVTLLFTYCLTFKRVSTWLRSDQPVQKTEPFIAY
IASKSMCTGLSVTRVLTAGHCVCSPMPYIKQVTLTLRNGDQGIHQPSGVK
APEWPSCTASRORRIRIOTLSGFALATAMLOMNSGRVIGLPPASDIPPGTGV
VFIVGYGRDNDRDSRRKNGGLTKKGRATIMECRHATNGNPICVYAGQNFQGLPAPGDS
LLPSQGVPLGVSHGVTLSPRLPIVYAVANMLDVRNSNI"
BASE COUNT 311 a 226 c 275 g 351 t
ORIGIN
Query Match 30.5%; Score 450.6; DB 3; Length 1163;
Best Local Similarity 83.9%; Pred. No. 4,1e-121;
Matches 608; Conservative 1; Mismatches 95; Indels 21; Gaps 8;
QY 1 GTGTCAACCTGGTGTGATACGTAGTGTGAACCTGTGCAACCGCACTGATTCGCCGTTT 60
DB 64 GTGTCAACCTGGTGTGATACGTAGTGTGAACCTGTGCAACCGCACTGATTCGCCGTTT 123
QY 61 ATTCGATTTTAAACAGAGAGAGAAATGTGTACAGSTTCACTAGTCTCAACGAGAGCA 120
DB 124 ATTCGATATATATGACATCAAGAAATCAATGTGTACAGSTTCACTAGTCTCAACGAGAGTA 183
QY 121 GACTACACAGCTGTGATTTGTTGCTCAACATTCGAGTGTGGGTAAGAGAT-CG 179
DB 184 GACTACACAGCTGTGATTTGTTGCTCAACATTCGAGTGTGGGTAAGAGATTAAC 243
QY 180 ACTGAAACATTTGTTG---CTCATATGATATGATTTGATTTCAACAGAGTGTTCG 236
DB 244 ACTGATATACATTTGTTGATGTAATGTAATGATTAATTTCAACAGAGT----- 294
QY 237 TCGGTGCGTTTGTGTGTGATGATGAGTGGGTTGATATGATTTATATTAATTA 296
DB 295 ---GTAACGTGTGTGTGTGTGATGAGTGGGTTGATTAATTAATGTAATTAATTA 350
QY 297 GGTCTATTTTCAAGTTTCACTTTCTCACAGGAGGATGAGCAACCAAGGATTCATCA 356
DB 351 TGTTC-CATTTTCAGTTTCACTTTCTCACATTTGAGGAGGAGCAACCAAGGATTCATCA 409
QY 357 CCACCGTCTGAGTTAGGTGGCACCGAGATACATGCCCTCTTGTATGTGGCAGACA 416

```

```

DB 410 TCAACCGTCTGAGTTAGGTAGGACCGAGATACATGCCCTCTTGTATGCGGTCACGACA 469
QY 417 GAGAGAGCAATTCGACAGACACTCAGTGATTCGATTTGCAATTTGATATGTCGCTCA 476
DB 470 GAGGAGAAATTCAGACAGACACTCAGTGATTTGCCATTTCMACTGAAATGCTGCTCA 529
QY 477 AATGTCACCTTACAGACTGTGAATCAGATGATCAGTGTGCGACAGCCATGATATCC 536
DB 530 AATGTCACCTTACAGACTGTGAATCAGATGATCAGTGTGCGACAGCCATGATATCC 589
QY 537 GCCACCTGAACTGTGTTCATTGTTGTTTATGGAAGGATGATTAACGACCGTGATCC 596
DB 590 GACACCTGAACTGTGTTCATTGTTGTTTATGGAAGGATGATTAACGACCGTGATCC 649
QY 597 GTACGTAAGATGGTGTGAATATTAAGAAAGTGTGTGTGTAATTAACGACATGAC 656
DB 650 GTCCGTAGAGCTGTGG-ATTTCAGAAAGGTAGATTTGTGTGTCACAAACACATAC 708
QY 657 TCAGTC-CTCAGTCAGAT-CTCAGTCAGTTATTTGTGTGTCTGTGATCTGTTGTG 714
DB 709 TCAGTCAGTCAGTCAGTTAGTAAGACAGTTAATCTGTCTATTTGTTATGTATCGGTT 768
QY 715 TCTGT 719
DB 769 GTTGT 773

RESULT 4
SCMCPROT 1152 bp mRNA linear INV 26-APR-1993
LOCUS S_mansoni cercarial protease mRNA, complete cds.
DEFINITION J03946
ACCESSION J03946.1 GI:160947
VERSION 1
KEYWORDS cercarial protease; serine protease.
SOURCE S_mansoni (Strain Puerto Rican) cercaria acetabular gland, CDNA to mRNA, clone B.
ORGANISM Schistosoma mansoni
Eukaryota; Metazoa; Platyhelminthes; Trematoda; Digenea;
Strigoidida; Schistosomatidae; Schistosomatidae; Schistosoma.
REFERENCE
1 (bases 1 to 1152)
Newport,G.R., McKerrow,J.H., Hedstrom,R., Pettit,M., McGarrigle,L.,
Barr,P.J. and Agabian,N.
Cloning of the proteinase that facilitates infection by schistosome
parasites
J. Biol. Chem. 263 (26), 13179-13184 (1988)
JOURNAL MEDLINE 88330818
PUBMED 316457
COMMENT Draft entry and clean copy of sequence [1] kindly submitted by
J.H.McKerrow, 13-JUN-1988.
FEATURES
source
1..1152
/organism="Schistosoma mansoni"
/db_xref="taxon:6183"
255..1049
/note="cercarial protease precursor"
/codon_start=1
/protein_id="AA29864.1"
/db_xref="GI:160948"
/translation="MSNMRFLVVTLLFTYCLTFKRVSTWLRSDQPVQKTEPFIAY
FLTERMCTGLSVTRVLTAGHCVCSPMPYIKQVTLTLRNGDQGIHQPSGVKVA
PGWPSGMSARORRIRIOTLSGFALATAMLOMNSGRVIGLPPASDIPPGTGV
FIVGYGRDNDRDSRRKNGGLTKKGRATIMECRHATNGNPICVYAGQNFQGLPAPGDS
GGPLPSIQGVPLGVSHGVTLSPRLPIVYAVANMLDVRNSNI"
255..335
sig_peptide
/note="cercarial protease signal peptide"
336..1046
mat_peptide
/product="cercarial protease"
BASE COUNT 345 a 273 c 230 g 304 t
ORIGIN
Query Match 22.5%; Score 332.4; DB 3; Length 1152;
Best Local Similarity 77.1%; Pred. No. 2.5e-86;

```

		Query Match	Best Local Similarity	Matches 301, Conservative	Score 287.2; DB 3;	Length 904;
source	gene	CDS				
1..904	/organism="Schistosoma mansoni"					
/db_xref="taxon:6183"						
1..904	/gene="CE-2a"					
/note="SmcE-2a"						
1..792	/gene="CE-2a"					
/note="protease"						
/codon_start=1						
/product="elastase 2a"						
/protein_id="PAAAM3941.1"						
/db_xref="GI:21217531"						
/translation="MLNGRTFLMVTLTFTYCLTEERVSTWLVRKEPVQDTREPYIAFVTRTCTGSLSTRALVTLATGHCVCSPMVVIVSPLTNNGDQGTHQPSGVKAP EYMSCTASRRRIROTLGGFDLATYMLQMVLGSIGVISLPQASDIPTGETDVFI VVGGRDDNDRDRPERBRRAGLILKKGRAITVMCKSTTGNPICVOAAYVFQITLRPGDSG GPRLSRPQEPVLGVSHSVTLNRLDLVLYRAYARLVGFVSNI"						
polya_signal						
BASE COUNT	239 a	200 c	224 g	241 t		
ORIGIN						
Query Match	19.4%; Score 287.2; DB 3;					
Best Local Similarity	92.9%; Pred. No. 4.8e-73;					
Matches 301, Conservative	0; Mismatches 23; Indels 0; Gaps 0;					
OY	305 TTCCAGTTTCATTTCACACTCAGGAATGGCGAACCAAGGCATCATCACCAACCGT	364				
DB	227 TTCAGGTTCATTTCACACTGAGAAATGGCGAACCAAGGCATCATCACCAACCGT	286				
OY	365 CTGAGATTAAAGGTGCCACCAGATPATCATGCCCTCTTGATGTCGSCACGACAGAGAAC	424				
DB	287 CTGAGATTAAAGGTGCCACCAGATPATCATGCCCTCTTGATGTCGSCACGACAGAGA	346				
OY	425 CAATGCCACAGACTCAGTGGATTGATTAATGCAATTGTAATGCTGCTCAAATGCTCA	484				
DB	347 GAATGAGCAGACACTCAGTGGATTGCAATTTGCAAACTGTATCTCGCTCAAAATGCTCA	406				
OY	485 ACTTACAGAGTGAATCAGAGTAGATCACTGTCGCCACAGCCATCGGAATATCCC GCCACCTG	544				
DB	407 ACTTACAAGTGAATCAGAGTAGATCACTGTCGCCACAGCCATCGGAATATCCCAGACCTG	466				
OY	545 GAACGTGCTGTTTTCTCATTTGTTGTTATGGAAGGATGATAAGCACTGATCCGTCACGTA	604				
DB	467 GAACGTGATGTTTTCTCATTTGTTGTTATGGAAGGATGATAAGCACTGATCCGTCGCTA	526				
OY	605 AGAATGTGGAATATTGGAAGAAAG	628				
DB	527 GAGCTGTGGGATTTTGAAGAAAG	550				
RESULT 6						
LOCUS	AF510341	943 bp	mRNA	linear	INV 01-JUN-2002	
DEFINITION	Schistosoma haematobium elastase 1a (CE-1a)			mRNA,	complete cds.	
ACCESSION	AF510341					
VERSION	AF510341.1	GI:21217534				
KEYWORDS	Schistosoma haematobium, Schistosoma haematobium Euharotoia; Metazoa; Platyhelminthes; Trematoda; Digenea; Strigidae; Schistosomatida; Schistosomidae; Schistosomidae; Schistosoma. 1 (bases 1 to 943) Salter,J.P., Choe,Y., Albrecht,H., Franklin,C., Lim,K.C., Craig,K.S. and McKerrow,J.H. Ceacarial elastase is encoded by a functionally conserved gene family across multiple species of schistosomes J. Biol. Chem. 277 (27), 24618-24624 (2002)					
ORGANISM	JOURNAL	PUBLISHED	22086216			
AUTHORS	MEDLINE	PUBMED	11986325			
REFERENCE	2 (bases 1 to 943)					
SEQUENCE	Salter,J.P., Choe,Y., Albrecht,H., Franklin,C., Lim,K.-C., Craig,K.C.					

TITLE and McKerrow, J.H.  
JOURNAL Direct Submission  
Submitted (08-MAY-2002) Pathology, University of California - San Francisco, 513 Parnassus, Box 0511, San Francisco, CA 94143, USA

FEATURES  
source Location/Qualifiers

1..943  
/organism="Schistosoma haematobium"  
/db\_xref="taxon:6185"  
gene 1..943  
/gene="CE-1a"  
/note="ShCE-1a"  
5'UTR 1..17  
/gene="CE-1a"  
18..809  
/gene="CE-1a"  
/note="protease"  
CDS /codon\_start=1  
/product="elastase 1a"  
/protein\_id="AA043943.1"  
/db\_xref="GI:21217535"

3'UTR 810..943  
/gene="CE-1a"  
polyA\_signal 919..924  
/gene="CE-1a"  
BASE COUNT 263 a 220 g 250 t  
ORIGIN

Query Match 18.9%; Score 279.2; DB 3; Length 943;  
Best Local Similarity 91.4%; Pred. No. 1.1e-70;  
Matches 296; Conservative 0; Mismatches 28; Indels 0; Gaps 0;

QY 305 TTCAGGTTTCATTTCTACACTGAGGAATGGCGACCAAGGAGCATCCATCAACCGCT 364  
DB 244 TTGGGTTTCATTTCTACACTGAGGAATGGCGACCAAGGAGCATCCATCAACCGCT 303  
QY 365 CTGAGTTAAGTGGGACACAGATATATGCTTGTATGTCGGACGACAGAGAGAC 424  
DB 304 CTGAGTGTGTGAGACACAGATATATGCTTGTATGTCGGACGACAGAGAGAC 363  
QY 425 CAATGCACGACACTCAGTGATGATGATGATGATGATGATGATGATGATGATGATGAT 484  
DB 364 CAATGCACGACACTCAGTGATGATGATGATGATGATGATGATGATGATGATGATGAT 423  
QY 485 ACTTACGAGTGAATCAGAGTGTGCTGCGACAGCCATGATCCGCGACCTG 544  
DB 424 ACTTACGAGTGAATCAGAGTGTGCTGCGACAGCCATGATCCGCGACCTG 483  
QY 545 GAACTGTGTGTTTTCATTTGTTGTTATGGAAGGATGATAACGACCGTGTCCGTCACTA 604  
DB 484 GAACTGTGTGTTTTCATTTGTTGTTATGGAAGGATGATAACGACCGTGTCCGTCACTA 543  
QY 605 AGAATGTGGAATATGAAGAAG 628  
DB 544 GAAATGTGGAATATGAAGAAG 567

RESULT 7 AFS10342 419 bp mRNA linear INV 01-JUL-2002  
AFS10342  
LOCUS Schistosoma haematobium elastase 1b (CE-1b) mRNA, partial cds.  
DEFINITION AFS10342  
ACCESSION AFS10342  
VERSION AFS10342.1 GI:21217536  
KEYWORDS  
SOURCE Schistosoma haematobium.  
ORGANISM Schistosoma haematobium.  
Eukaryota; Metazoa; Platyhelminthes; Trematoda; Digenea;  
Strigeidida; Schistosomatoidae; Schistosomatidae; Schistosoma.  
REFERENCE 1 (bases 1 to 419)

AUTHORS Salter, J.P., Choe, Y., Albrecht, H., Franklin, C., Lim, K.C.,  
Craik, C.S. and McKerrow, J.H.  
TITLE Cercarial elastase is encoded by a functionally conserved gene  
JOURNAL family across multiple species of schistosomes  
J. Biol. Chem. 277 (27), 24618-24624 (2002)

MEDLINE 22086216

PUBMED 11986325

REFERENCE 2 (bases 1 to 419)

AUTHORS Salter, J.P., Choe, Y., Albrecht, H., Franklin, C., Lim, K.-C., Craik, C.

and McKerrow, J.H.

TITLE Direct Submission

JOURNAL Submitted (08-MAY-2002) Pathology, University of California - San

Francisco, 513 Parnassus, Box 0511, San Francisco, CA 94143, USA

FEATURES  
source Location/Qualifiers

1..419  
/organism="Schistosoma haematobium"  
/db\_xref="taxon:6185"  
gene <1..>419  
/gene="CE-1b"  
/note="ShCE-1b"  
CDS <1..>419  
/gene="CE-1b"  
/note="protease"  
/codon\_start=1  
/product="elastase 1b"  
/protein\_id="AA043944.1"  
/db\_xref="GI:21217537"

3'UTR 419..419  
/gene="CE-1b"  
polyA\_signal 419..419  
/gene="CE-1b"  
BASE COUNT 118 a 104 c 102 g 95 t  
ORIGIN

Query Match 16.8%; Score 248.8; DB 3; Length 419;  
Best Local Similarity 90.8%; Pred. No. 9.1e-62;  
Matches 265; Conservative 0; Mismatches 27; Indels 0; Gaps 0;

QY 337 GACCAACAAGGATCATCATCAACCAACCGTCTGGAGTTAAGTGGACCAAGATACATGCC 396  
DB 1 GACCAACAAGGATCATCATCAACCAACCGTCTGGAGTTAAGTGGACCAAGATACATGCC 60  
QY 397 TTTTGTATGTGGGACGACAGAGAGAGACCAATCCGACAGACACTAGTGGATTCATATT 456  
DB 61 TTTTGTATGTGGGACGACAGAGAGAGACCAATCCGACAGACACTAGTGGATTCATATT 120  
QY 457 GCAATTTGATGCTGGCTCAATAGTGTCAACTTACAGAGTGAATCAGATGATCAGTCTG 516  
DB 121 GCAATTTGATGCTGGCTCAATAGTGTCAACTTACAGAGTGAATCAGATGATCAGTCTG 180  
QY 517 CCACAGCCATCGGATATCCCGCACCTGGAACTGGTGTTCATTTGTTATGGAAG 576  
DB 181 CCACAGCCATCGGATATCCCGCACCTGGAACTGGTGTTCATTTGTTATGGAAG 240  
QY 577 GATGATTAACGACCGTGTATCCGTCACTGAAGAATGTGGAATTTGAAGAAAG 628  
DB 241 GATGATTAACGACCGTGTATCCGTCACTGAAGAATGTGGAATTTGAAGAAAG 292

RESULT 8 AFS10343 431 bp mRNA linear INV 01-JUL-2002  
AFS10343  
LOCUS Schistosomium douthitti elastase 1a (CE-1a) mRNA, partial cds.  
DEFINITION AFS10343  
ACCESSION AFS10343  
VERSION AFS10343.1 GI:21217538  
KEYWORDS  
SOURCE Schistosomium douthitti.  
ORGANISM Schistosomium douthitti.  
Eukaryota; Metazoa; Platyhelminthes; Trematoda; Digenea;  
Strigeidida; Schistosomatoidae; Schistosomatidae; Schistosomium.  
REFERENCE 1 (bases 1 to 431)  
AUTHORS Salter, J.P., Choe, Y., Albrecht, H., Franklin, C., Lim, K.C.,  
Craik, C.S. and McKerrow, J.H.  
TITLE Cercarial elastase is encoded by a functionally conserved gene

family across multiple species of schistosomes  
JOURNAL J. Biol. Chem. 277 (27), 24618-24624 (2002)  
MEDLINE 22086216  
PUBMED 11986325  
REFERENCE 2 (bases 1 to 431)  
AUTHORS Salter, J.P., Choe, Y., Albrecht, H., Franklin, C., Lim, K.-C., Craik, C. and McKerrow, J.H.  
TITLE Direct Submission  
JOURNAL Submitted (08-MAY-2002) Pathology, University of California - San Francisco, 513 Parnassus, Box 0511, San Francisco, CA 94143, USA  
FEATURES  
source location/Qualifiers  
1..431  
/organism="Schistosoma mansoni"  
/db\_xref="taxon:38741"  
gene /gene="CE-1a"  
/note="SdCE-1a"  
<1..>431  
<1..>431  
CDS /gene="CE-1a"  
/note="protease"  
/codon\_start=1  
/product="elastase 1a"  
/protein\_id="AAM43945.1"  
/db\_xref="GI:21217539"  
translation="DQGGIHQPPSGVVAPEYMPSCASSRQRVYKQTLGFDIAIVL  
LAEMVNLQGIKIVSLPQPTDIPAGTPEVFIYGRDNDNDPSRRNGILKKGRATV  
LAEKSTTGNPIGVGRYVFGQIPAPGDSGGELPSPQNSIS"  
BASE COUNT 117 a 108 c 109 g 97 t  
ORIGIN  
Query Match 16.3%; Score 240.8; DB 3; Length 431;  
Best Local Similarity 89.0%; Pred. No. 2.1e-59;  
Matches 260; Conservative 0; Mismatches 32; Indels 0; Gaps 0;  
QY 337 GACCAACAAAGCATTCATACCAACCGCTTGAGTTAAGTGGCCACGAGATACATCC 396  
DB 1 GACCAACAAAGCATTCATACCAACCGCTTGAGTTAAGTGGCCACGAGATACATCC 60  
QY 397 TTTGTAATGTCGACGAGAGAGACCAATCCGACAGACTCATGTGATTCGATATT 456  
DB 61 TTTGTAATGTCGACGAGAGAGACCAATCCGACAGACTCATGTGATTCGATATT 120  
QY 457 GCATTTGTAATGTCGACGAGAGAGACCAATCCGACAGACTCATGTGATTCGATATT 516  
DB 121 GCATTTGTAATGTCGACGAGAGAGACCAATCCGACAGACTCATGTGATTCGATATT 180  
QY 517 CCACAGCCATCGGATATCCCGCACCCTGGAAGTGTGTTTCATTTGTTGATTAAGAG 576  
DB 181 CCACAGCCATCGGATATCCCGCACCCTGGAAGTGTGTTTCATTTGTTGATTAAGAG 240  
QY 577 GATGATTAACGACCGTATCCGTCACCTAGAAATGTTGAAGAAG 628  
DB 241 GATGATTAACGACCGTATCCGTCACCTAGAAATGTTGAAGAAG 292

RESULT 9  
AF510340 886 bp mRNA linear INV 01-JUL-2002  
DEFINITION Schistosoma mansoni elastase 2b (CE-2b) mRNA, complete cds.  
ACCESSION AF510340  
VERSION AF510340.1 GI:21217532  
KEYWORDS  
SOURCE Schistosoma mansoni.  
ORGANISM Schistosoma mansoni  
Eukaryota; Metazoa; Platyhelminthes; Trematoda; Digenea; Strigeidae; Schistosomatidae; Schistosomatidae; Schistosoma.  
REFERENCE 1 (bases 1 to 886)  
AUTHORS Salter, J.P., Choe, Y., Albrecht, H., Franklin, C., Lim, K.-C., Craik, C.S. and McKerrow, J.H.  
TITLE Cercarial elastase is encoded by a functionally conserved gene family across multiple species of schistosomes  
JOURNAL J. Biol. Chem. 277 (27), 24618-24624 (2002)  
MEDLINE 22086216

PUBMED 11986325  
REFERENCE 2 (bases 1 to 886)  
AUTHORS Salter, J.P., Choe, Y., Albrecht, H., Franklin, C., Lim, K.-C., Craik, C. and McKerrow, J.H.  
TITLE Direct Submission  
JOURNAL Submitted (08-MAY-2002) Pathology, University of California - San Francisco, 513 Parnassus, Box 0511, San Francisco, CA 94143, USA  
FEATURES  
source location/Qualifiers  
1..886  
/organism="Schistosoma mansoni"  
/db\_xref="taxon:6183"  
gene /gene="CE-2b"  
/note="smCE-2b"  
1..792  
CDS /gene="CE-2b"  
/note="protease"  
/codon\_start=1  
/product="elastase 2b"  
/protein\_id="AAM43942.1"  
/db\_xref="GI:21217533"  
translation="MNGWTFILVTLFTYVILTCQVSTWILRSGPVOQRTEPPIAL  
LMDASHCTGSLVSSRAVLTRGHCVCGQPIVRVSTLSFDDRTINRPLEIKVAP  
EYNPVQQLKRENKRITKSLGVDMAIITLNLVLEFGVKVTSIAAEIDIPISIAV  
MVGQDIDRDPPSGRGYGIKKKSATIMACRHRHTFDPICVKSGPNSKOIAGPDSG  
GPLLITPOGPIVGAANGVFLPALADLCVERSVPRMLKFLIPNI"  
polyA\_signal 872..877  
BASE COUNT 257 a 165 c 191 g 273 t  
ORIGIN  
Query Match 9.7%; Score 143.4; DB 3; Length 886;  
Best Local Similarity 64.7%; Pred. No. 1e-30;  
Matches 213; Conservative 0; Mismatches 116; Indels 0; Gaps 0;  
QY 305 TTCAGATTTCATTTCTCACTGAGAAATGGGACCAAGGATTCATCAACCAACGCT 364  
DB 227 TTCAGATTTCATTTCTCACTGAGAAATGGGACCAAGGATTCATCAACCAACGCT 286  
QY 365 CTGAGTTAAGTGGCACCAGATACATGCCCTCTTGATATGTCGACAGAGAGAGAC 424  
DB 287 TGGAAATTAAAGTTGACACAGATTAATCCGTGTCAGCTGAACGAGAAATTAAC 346  
QY 425 CAATGCAACAGACTCATGTGATTCGATTTGCAATTGTAATGCTGCTCAATATGTC 484  
DB 347 GCATTACGAGTCACTTGGTGTATGATATGCAATTATACACTCAATATTGTC 406  
QY 485 ACTTACAGAGTGAATCAGAGTATCAGTCTGCCACACCAATCGGATATCCGCCACCTG 544  
DB 407 ACTTACAGAGTGAATCAGAGTATCAGTCTGCCACACCAATCGGATATCCGCCACCTG 466  
QY 545 GAACGTGTTTTCATTTGTTGTTATGGAAGGATGATAGACCGTATCCGTACGTA 604  
DB 467 AAAGTATTGCTTATATGATGATGATGATGATGATGATGATGATGATGATGATGAT 526  
QY 605 AGAATGTTGAATATTGAAGAAGTGAAGT 633  
DB 527 GATATGTTGAATTTTGAAGAAAGTGAAGT 555

RESULT 10  
AF510344 441 bp DNA linear INV 01-JUL-2002  
DEFINITION Schistosoma douthetti elastase 1b (CE-1b) gene, partial cde.  
ACCESSION AF510344  
VERSION AF510344.1 GI:21217540  
KEYWORDS  
SOURCE Schistosoma douthetti.  
ORGANISM Schistosoma douthetti  
Eukaryota; Metazoa; Platyhelminthes; Trematoda; Digenea; Strigeidae; Schistosomatidae; Schistosomatidae; Schistosoma.  
REFERENCE 1 (bases 1 to 441)  
AUTHORS Salter, J.P., Choe, Y., Albrecht, H., Franklin, C., Lim, K.-C.,





----- Project Information	
Center project name:	TURT
Center clone name:	CH230-2F11
----- Summary Statistics	
Sequencing vector:	Plasmid;
Chemistry:	Dye-terminator Big Dye; 100% of reads
Assembly program:	Phrap; version 0.990329
Consensus quality:	130648 bases at least Q40
Consensus quality:	139589 bases at least Q30
Consensus quality:	146680 bases at least Q20
-----	
* NOTE: Estimated insert size may differ from sequence length.	
* (see <a href="http://www.hsbc.bcm.tmc.edu/docs/genbank_draft_data.html">http://www.hsbc.bcm.tmc.edu/docs/genbank_draft_data.html</a> )	
* NOTE: This is a "working draft" sequence. It currently	
* consists of 80 contigs. The true order of the pieces	
* is not known and their order in this sequence record is	
* arbitrary. Gaps between the contigs are represented as	
* runs of N, but the exact sizes of the gaps are unknown.	
* This record will be updated with the finished sequence	
* as soon as it is available and the accession number will	
* be preserved.	
1	1641: contig of 1641 bp in length
* 1642	1741: gap of unknown length
* 1742	2841: contig of 1100 bp in length
* 2842	2941: gap of unknown length
* 2942	4280: contig of 1339 bp in length
* 4281	4380: gap of unknown length
* 4381	5820: contig of 1440 bp in length
* 5821	5920: gap of unknown length
* 5921	7065: contig of 1146 bp in length
* 7067	7166: gap of unknown length
* 7167	8688: contig of 1522 bp in length
* 8689	8788: gap of unknown length
* 8789	10123: contig of 1335 bp in length
* 10124	10223: gap of unknown length
* 10224	11365: contig of 1143 bp in length
* 11367	11466: gap of unknown length
* 11467	12779: contig of 1313 bp in length
* 12780	12879: gap of unknown length
* 12880	14430: contig of 1551 bp in length
* 14431	14530: gap of unknown length
* 14531	15911: contig of 1381 bp in length
* 15912	16011: gap of unknown length
* 16012	17548: contig of 1537 bp in length
* 17549	17648: gap of unknown length
* 17649	18792: contig of 1144 bp in length
* 18793	18892: gap of unknown length
* 18893	20263: contig of 1377 bp in length
* 20270	20369: gap of unknown length
* 20370	21467: contig of 1098 bp in length
* 21468	21567: gap of unknown length
* 21568	22808: contig of 1241 bp in length
* 22809	22908: gap of unknown length
* 22909	24735: contig of 1627 bp in length
* 24736	24835: gap of unknown length
* 24836	25859: contig of 1024 bp in length
* 25860	25959: gap of unknown length
* 25960	27399: contig of 1440 bp in length
* 27400	27499: gap of unknown length
* 27500	28743: contig of 1244 bp in length
* 28744	28843: gap of unknown length
* 28844	30648: contig of 1805 bp in length
* 30649	30748: gap of unknown length
* 30749	32159: contig of 1411 bp in length
* 32160	32259: gap of unknown length
* 32260	34628: contig of 2369 bp in length
* 34629	34728: gap of unknown length
* 34729	36607: contig of 1879 bp in length
* 36608	36707: gap of unknown length
* 36708	38242: contig of 1535 bp in length
* 38243	38342: gap of unknown length
* 38343	40307: contig of 1965 bp in length
* 40308	40407: gap of unknown length

*	40508	42404:	contig of 1997 bp in length
*	42405	42504:	gap of unknown length
*	42505	43873:	contig of 1369 bp in length
*	43874	43973:	gap of unknown length
*	43974	45267:	contig of 1294 bp in length
*	45268	45367:	gap of unknown length
*	45368	46430:	contig of 1063 bp in length
*	46431	46530:	gap of unknown length
*	46531	47774:	contig of 1244 bp in length
*	47775	47874:	gap of unknown length
*	47875	49048:	contig of 1174 bp in length
*	49049	49148:	gap of unknown length
*	49149	51961:	contig of 2813 bp in length
*	51962	52061:	gap of unknown length
*	52062	53527:	contig of 1466 bp in length
*	53528	53627:	gap of unknown length
*	53628	55506:	contig of 1879 bp in length
*	55507	55606:	gap of unknown length
*	55607	58069:	contig of 2463 bp in length
*	58070	58169:	gap of unknown length
*	58170	59986:	contig of 1829 bp in length
*	59989	60098:	gap of unknown length
*	60099	61741:	contig of 1643 bp in length
*	61742	61841:	gap of unknown length
*	61842	63804:	contig of 1963 bp in length
*	63805	63904:	gap of unknown length
*	63905	65921:	contig of 2017 bp in length
*	65922	66021:	gap of unknown length
*	66022	67937:	contig of 1916 bp in length
*	67938	68037:	gap of unknown length
*	68038	69385:	contig of 1348 bp in length
*	69386	70485:	gap of unknown length
*	69486	70741:	contig of 1256 bp in length
*	70742	70841:	gap of unknown length
*	70842	72409:	contig of 1568 bp in length
*	72410	72509:	gap of unknown length
*	72510	74686:	contig of 2177 bp in length
*	74687	74786:	gap of unknown length
*	74787	76424:	contig of 1638 bp in length
*	76425	76524:	gap of unknown length
*	76525	79032:	contig of 2508 bp in length
*	79133	79132:	gap of unknown length
*	79133	80840:	contig of 1708 bp in length
*	80841	80940:	gap of unknown length
*	80941	83491:	contig of 2551 bp in length
*	83492	83591:	gap of unknown length
*	83592	85875:	contig of 2284 bp in length
*	85876	85975:	gap of unknown length
*	85976	87712:	contig of 1737 bp in length
*	87713	87812:	gap of unknown length
*	87813	89415:	contig of 1603 bp in length
*	89416	89515:	gap of unknown length
*	89516	93192:	contig of 3677 bp in length
*	93193	93292:	gap of unknown length
*	93293	94976:	contig of 1684 bp in length

[illegible]



```

For further details see http://www.chori.org/bacpac/home.htm
VECTOR: pbac3.6.
Location/Qualifiers
FEATURES
    source
        1..199649
            /organism="Mus musculus"
            /db_xref="taxon:10090"
            /chromosome="11"
            /clone="RP23-327K3"
            /clone_1kb="RPC1-23"
BASE COUNT      54761 a 42696 c 44329 g 57863 t
ORIGIN
Query Match      4.2%; Score 61.4; DB 10; Length 199649;
Bseq Local Similarity 49.9%; Pred. No. 2.1e-06;
Matches 270; Conservative 0; Mismatches 256; Indels 15; Gaps 4;
QY 668 TCAGATGTCAGTCAAGTATTTCTGTGTGTCGTGTATCTGTTTGTCGTCTCTGTCTAC 727
Db 73858 TTTAAATGTTTGTGTGTGTGATGTGTATGTATGTATGTATGTATGTATGTATGTGTGT 73917
QY 728 CTGATCCGGTGTGTATATTGTGTACAGCGCTGTATATACAACTGTGTTGGATGACTTT 787
Db 73918 ATGCGTGTGTATGTAT--GTGTGTGTGTGTATGTGTGTGTGTGTGTGTGTGTGTGTG 73975
QY 788 GTGACAGTTCAAGTACAGAGTGAATTTCCATCTCGTCACTTGTGTGTGTGTGAGGTAGAGTGA 847

```

[illegible]

DB	74327	GTGTCGTCTGCATCATGTGTCGTGTCGTATGCGTACGTCGTGTGTGTCGTGTTCTCACTT	74386
QY	1204	C 1204	
Db	74387	C 74387	
RESULT 14			
AL772338/c			
LOCUS			
DEFINITION	AL772338	236562 bp DNA linear HTG 17-AUG-2002	
ACCSSION	Mus musculus chromosome 4 clone RP23-47A6, *** SEQUENCING IN PROGRESS ***, 4 unordered pieces.		
VERSION	AL772338		
KEYWORDS	AL772338.2 GI:22416014		
SOURCE	HTG; HTGS_PHASE1; HTGS_ACTIVEFIN; HTGS_DRAFT; HTGS_FULLTOP. house mouse.		
ORGANISM	Mus musculus Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus. 1 (bases 1 to 236562)		
REFERENCE	Hall, R.		
AUTHORS	Direct Submission		
TITLE	Submitted (16-AUG-2002) Wellcome Trust Sanger Institute, Hinxton,		
JOURNAL			







```

* 16001 16100: gap of unknown length
* 16101 17585: contig of 1485 bp in length
* 17585 17685: gap of unknown length
* 17685 19311: contig of 1626 bp in length
* 19311 19412: gap of unknown length
* 19412 20973: contig of 1562 bp in length
* 20973 21073: gap of unknown length
* 21073 22764: contig of 1691 bp in length
* 22764 22864: gap of unknown length
* 22864 24587: contig of 1723 bp in length
* 24587 24687: gap of unknown length
* 24687 26188: contig of 1501 bp in length
* 26188 26288: gap of unknown length
* 26288 27412: contig of 1124 bp in length
* 27412 28633: contig of 1121 bp in length
* 28633 28733: gap of unknown length
* 28733 30312: contig of 1579 bp in length
* 30312 30412: gap of unknown length
* 30412 32106: contig of 1694 bp in length
* 32106 34135: contig of 1929 bp in length
* 34135 34235: gap of unknown length
* 34235 36177: contig of 1942 bp in length
* 36177 36278: gap of unknown length
* 36278 38382: contig of 2105 bp in length
* 38382 40989: gap of unknown length
* 40989 41089: contig of 2507 bp in length
* 41089 43198: gap of unknown length
* 43198 43298: gap of unknown length
* 43298 46975: contig of 3677 bp in length
* 46975 47075: gap of unknown length
* 47075 48596: contig of 1521 bp in length
* 48596 48696: gap of unknown length
* 48696 53355: contig of 4659 bp in length
* 53355 53455: gap of unknown length
* 53455 57098: contig of 3643 bp in length
* 57098 57199: gap of unknown length
* 57199 59834: contig of 2636 bp in length
* 59834 59934: gap of unknown length
* 59934 64544: contig of 4610 bp in length
* 64544 64644: gap of unknown length
* 64644 70344: contig of 5700 bp in length
* 70344 70444: gap of unknown length
* 70444 74027: contig of 3583 bp in length
* 74027 74127: gap of unknown length
* 74127 79625: contig of 5498 bp in length
* 79625 79725: gap of unknown length
* 79725 84722: contig of 4997 bp in length
* 84722 84822: gap of unknown length
* 84822 89618: contig of 4796 bp in length
* 89618 89718: gap of unknown length
* 89718 95248: contig of 5530 bp in length
* 95248 95349: gap of unknown length
* 95349 101835: contig of 6487 bp in length
* 101835 101935: gap of unknown length
* 101935 106434: contig of 4499 bp in length
* 106434 106534: gap of unknown length
* 106534 113798: contig of 7264 bp in length
* 113798 113898: gap of unknown length
* 113898 118895: contig of 4997 bp in length
* 118895 118996: gap of unknown length
* 118996 125057: contig of 6062 bp in length
* 125057 125157: gap of unknown length
* 125157 134617: contig of 9460 bp in length
* 134617 134717: gap of unknown length
* 134717 143217: contig of 8500 bp in length
* 143217 143317: gap of unknown length
* 143317 151457: contig of 8140 bp in length
* 151457 151557: gap of unknown length
* 151557 158975: contig of 7418 bp in length
* 158975 159075: gap of unknown length

```

```

* 159076 167023: contig of 7948 bp in length
* 167023 167123: gap of unknown length
* 167123 167124: contig of 10325 bp in length
* 167124 177548: gap of unknown length
* 177548 177549: gap of unknown length
* 177549 190566: contig of 13018 bp in length
* 190566 190667: gap of unknown length
* 190667 202789: contig of 12123 bp in length
* 202789 202890: gap of unknown length
* 202890 218636: contig of 15747 bp in length
* 218636 218737: gap of unknown length
* 218737 248706: contig of 29970 bp in length.

FEATURES
    source                1..248706

Query Match
Best Local Similarity 35.2%; Pred. No. 1.3e-05;
Matches 193; Conservative 0; Mismatches 352; Indels 3; Gaps 1;

Qy 678 GTCAGTATTCGTGCTGTGTGATCTCTGTTGCTGCTGTCTGATCCGGT 737
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 37411 GATGTGAGTGTGTGTGATGATGAATGATGATGATGATGATGATGATG 37470

Qy 728 TGTGTATTGTCAGAGCCTTGATATACAACTGTGTTGGATGACTTTGTGACAGTTC 797
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 37471 GAGAGTGTGTGTGATATATGATATGATGATGATGATGATGATGATGATG 37530

Qy 798 AGTACAGAGCATTTCCATCTCGGTGATGTTGTTGAGAGTGAAGTACGTATGTA 857
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 37531 GTGACTATATGAGAAATGATGTGTGTGAATGTGTGTGAGTACGATCTGAGAGT 37590

Qy 858 GGTGAGTTGAGGTAGATGATGATGAGTGAATGTGATGATGATGATGATGATG 917
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 37591 ATATAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 37647

Qy 918 TCGAGAGAGAACTCATGAAATATCTATCAAAAGTGAAGTGTGTTGTATCATGA 977
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 37648 TATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 37707

Qy 978 AGTAGGGGTCATGTGTTGATGATGTTGAGAGTGTGAGATGATGAGATGATCTTG 1037
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 37708 NNNNNNTNNNNNTNNNNNTNNNNNTNNNNNTNNNNNTNNNNNTNNNNNTNNNN 37767

Qy 1038 ATCTGCAATATAGTACATGATGATGATGATGATGATGATGATGATGATGATG 1097
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 37768 NNTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT 37827

Qy 1098 GCGTCGATATGTGCACTGATATTTTGAATTCACCTGTGTGTTGTTGTTCCG 1157
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 37828 NNTTGTGTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT 37887

Qy 1158 TTGCTTTTGTCTTCTTCATCTGACGTGATGATGATGATGATGATGATGATGATG 1217
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 37888 TTTTTTTTTTCTCTCTNNNTNTNTTTTTTTTTTTTTTTTCTTTCTTNTTTTTTTTTTT 37947

Qy 1218 GTTGTGTT 1225
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 37948 TTTTTTTT 37955

RESULT 17
AC098634 105899 bp DNA linear HTG 12-JUN-2002
AC098634 Rattus norvegicus clone CH230-111G8,*** SEQUENCING IN PROGRESS
DEFINITION Rattus norvegicus clone CH230-111G8,***
ACCESSION AC098634
VERSION AC098634.4 GI:21729818
KEYWORDS HTG; HTGS PHASE1.
SOURCE Norway rat.
ORANISM Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
REFERENCE
1 (bases 1 to 105899)

```

## AUTHORS

Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C., Albrooks,S.L., Amarantunga,H.C., Are,J.R., Ayele,M., Banks,T., Barbra,J., Benton,J., Bimaga,K., Blankenburg,K., Bonnin,D., Bouck,J., Bowie,S., Brileva,M., Brown,E., Brown,M., Bryant,N.P., Bunay,C., Butch,P., Burkett,C., Burrell,K.L., Byrd,N.C., Caron,T.F., Carter,M., Cavazos,S.R., Chacko,O., Chavez,D., Chen,G., Chen,R., Chen,Z., Chowdhry,I., Christopoulos,C., Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R., Davila,M.L., Davis,C., Davy-Carroll,L., Dederich,D.A., Delaney,K.R., Delgado,O., Denn,A.L., Ding,Y., Dinh,H.H., Douthwaite,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.J., Earnhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Escotto,M., Falls,T., Ferraguto,D., Edwards,C.C., Elhaj,C., Escotto,M., Farris,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R., Gorrell,J.H., Guevara,M., Gunaratne,P., Hale,S., Hamilton,K., Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A., Hernandez,J., Hernandez,O., Hodgson,A., Hogues,M., Holloway,C., Hollins,B., Homel,F., Howard,S., Huber,J., Hulyk,S., Hume,J., Jackson,L.E., Jacobson,B., Jia,Y., Johnson,R., Jolivet,S., Joudah,S., Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J., Kovar,C., Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C., Lewis,L., Li,J., Li,Z., Lichteberg,O., Lieu,C., Liu,J., Liu,W., Louissege,H., Lofado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R., Ma,J., Maheshwari,M., Mapua,P., Martin,R., Martindale,A., Martinez,E., Massey,E., Mawhinney,E., McLeod,M.P., Meador,M., Mei,G., Metzker,M., Miner,G., Miner,Z., Mitchell,T., Mohabbat,K., Morgan,M., Morris,S., Moser,M., Neal,D., Newton,J., Newton,N., Nguyen,A., Nguyen,N., Nguyen,N., Nickerson,E., Nwokenkwo,S., Oguh,M., Okumu,G., Oragunye,N., Oviedo,R., Pace,A., Payton,B., Peery,J., Perez,L., Peters,L., Pickens,R., Pirmus,E., Pu,L.L., Qulles,M., Ren,Y., Rives,M., Rojas,A., Rojibokan,I., Rolfe,M., Ruiz,S., Savary,G., Scherer,S., Scott,G., Shen,H., Shooshitari,N., Sisson,I., Sodegren,E., Sonaiki,T., Sparks,A., Stanley,H., Stone,H., Sutton,A., Statek,A., Tabor,P., Tamerisa,A., Tamerisa,K., Tang,H., Tansey,J., Taylor,C., Taylor,T., Telitod,B., Thomas,N., Thomas,S., Tsumaki,K., Vasquez,L., Vera,V., Villalón,D., Vinson,R., Wang,Q., Wang,S., Ward-Moore,S., Warren,R., Washington,C., Watlington,S., Williams,G., Williamson,A., Wleczek,R., Wooden,S., Worley,K., Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D., Weinstein,G., and Gibbs,R.

## TITLE

## JOURNAL

## REFERENCE

## AUTHORS

## TITLE

## JOURNAL

## REFERENCE

## AUTHORS

## TITLE

## JOURNAL

## REFERENCE

## AUTHORS

## TITLE

## JOURNAL

## REFERENCE

## AUTHORS

## TITLE

## JOURNAL

## REFERENCE

## AUTHORS

## TITLE

## JOURNAL

## REFERENCE

## AUTHORS

## TITLE

## JOURNAL

## REFERENCE

## AUTHORS

## TITLE

## JOURNAL

## REFERENCE

## AUTHORS

## TITLE

## JOURNAL

## REFERENCE

## AUTHORS

## TITLE

## JOURNAL

## REFERENCE

\* NOTE: This is a 'working draft' sequence. It currently consists of 54 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as \* runs of N, but the exact sizes of the gaps are unknown. This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved.

1 1196: contig of 1196 bp in length  
\* 1197 1296: gap of unknown length  
\* 1297 1297: contig of 1250 bp in length  
\* 1297 2546: gap of unknown length  
\* 2547 2647: gap of 1296 bp in length  
\* 2647 3943: gap of unknown length  
\* 3943 4042: contig of 1302 bp in length  
\* 4042 4043: gap of unknown length  
\* 4043 5345: gap of 1247 bp in length  
\* 5345 5445: gap of unknown length  
\* 5445 6691: gap of 1216 bp in length  
\* 6691 6791: gap of unknown length  
\* 6791 8007: contig of 1471 bp in length  
\* 8007 8107: gap of unknown length  
\* 8107 8108: gap of 1176 bp in length  
\* 8108 9579: gap of unknown length  
\* 9579 9679: gap of unknown length  
\* 9679 10855: gap of 1511 bp in length  
\* 10855 10955: gap of unknown length  
\* 10955 12466: gap of 1749 bp in length  
\* 12466 12566: gap of unknown length  
\* 12566 14315: gap of 1054 bp in length  
\* 14315 14415: gap of unknown length  
\* 14415 14568: gap of unknown length  
\* 14568 15668: gap of 1614 bp in length  
\* 15668 15669: gap of unknown length  
\* 15669 17183: gap of 1345 bp in length  
\* 17183 17282: gap of unknown length  
\* 17282 18627: gap of unknown length  
\* 18627 18728: gap of 1063 bp in length  
\* 18728 19791: gap of unknown length  
\* 19791 19891: gap of 1365 bp in length  
\* 19891 21256: gap of unknown length  
\* 21256 21356: gap of 1373 bp in length  
\* 21356 22728: gap of unknown length  
\* 22728 24070: gap of 1242 bp in length  
\* 24070 24071: gap of unknown length  
\* 24071 24171: gap of 1062 bp in length  
\* 24171 25232: gap of unknown length  
\* 25232 25332: gap of 1267 bp in length  
\* 25332 26699: gap of unknown length  
\* 26699 26700: gap of 1603 bp in length  
\* 26700 28402: gap of unknown length  
\* 28402 28502: gap of 1997 bp in length  
\* 28502 30499: gap of unknown length  
\* 30499 30500: gap of 1726 bp in length  
\* 30500 32325: gap of unknown length  
\* 32325 32425: gap of 2208 bp in length  
\* 32425 34633: gap of unknown length  
\* 34633 34733: gap of unknown length  
\* 34733 36722: contig of 1989 bp in length  
\* 36722 36723: gap of unknown length  
\* 36723 38362: contig of 1540 bp in length  
\* 38362 38463: gap of unknown length  
\* 38463 40346: gap of 1884 bp in length  
\* 40346 40347: gap of unknown length  
\* 40347 42311: contig of 1664 bp in length  
\* 42311 42410: gap of unknown length  
\* 42410 44174: contig of 1764 bp in length  
\* 44174 44274: gap of unknown length  
\* 44274 46305: gap of 2031 bp in length  
\* 46305 46306: gap of unknown length  
\* 46306 47639: contig of 1234 bp in length  
\* 47639 47740: gap of unknown length  
\* 47740 49253: contig of 1414 bp in length  
\* 49253 51108: gap of unknown length  
\* 51108 51208: gap of 1855 bp in length  
\* 51208 53770: contig of 2562 bp in length

\* NOTE: Estimated insert size may differ from sequence length  
(see [http://www.hgsc.bcm.tmc.edu/docs/Genbank\\_draft\\_data.html](http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html)).

```

* 53771 53870: gap of unknown length
* 53871 56088: contig of 2218 bp in length
* 56089 56188: gap of unknown length
* 56189 57810: contig of 1622 bp in length
* 57811 60241: gap of unknown length
* 60242 60341: contig of 231 bp in length
* 60342 62349: gap of unknown length
* 62349 62350: contig of 1908 bp in length
* 62350 64990: gap of unknown length
* 64991 65090: gap of unknown length
* 65091 67909: contig of 2819 bp in length
* 67910 68009: gap of unknown length
* 68010 69851: contig of 1842 bp in length
* 69852 69952: gap of unknown length
* 69952 72034: contig of 2083 bp in length
* 72035 72134: gap of unknown length
* 72135 74656: contig of 2522 bp in length
* 74657 74756: gap of unknown length
* 74757 76348: contig of 1592 bp in length
* 76349 76449: gap of unknown length
* 76449 79251: contig of 2803 bp in length
* 79252 79351: gap of unknown length
* 79352 80707: contig of 1356 bp in length
* 80708 80807: gap of unknown length
* 80808 82918: contig of 2111 bp in length
* 82919 83019: gap of unknown length
* 83019 85401: contig of 2383 bp in length
* 85402 85501: gap of unknown length
* 85502 88054: contig of 2553 bp in length
* 88055 88154: gap of unknown length
* 88155 90296: contig of 2142 bp in length
* 90297 90396: gap of unknown length
* 90397 92886: contig of 2490 bp in length
* 92887 92986: gap of unknown length
* 92987 96055: contig of 3069 bp in length
* 96056 96155: gap of unknown length
* 96156 98187: contig of 2032 bp in length
* 98188 98287: gap of unknown length
* 98288 101171: contig of 2884 bp in length
* 101172 101271: gap of unknown length
* 101272 105899: contig of 4628 bp in length.

Query Match      3.9%  Score 57.8;  DB 2;  Length 105899;
Best Local Similarity 36.5%  Pred. No. 2.3e-05;
Matches 203;  Conservative 0;  Mismatches 351;  Indels 2;  Gaps 1;

```

```

Db 18413 GTGGGNGTNGNGGTGGTGTGTTTGTGTTTGTGTTNNNTNTTTTGGGGTG 18412
Qy 1091 GTGTAAAGGCGATATATGTCACGATGATATTTGCAATCAGTGTGTTGTTT 1150
Db 18473 GTGTTNNGTTNGGTTTGGGTTNNNNGTTTNTTGTNNNTTGTGANNNTTGN 18532
Qy 1151 TGTTCGTTGTTGTTTGTCTTTCATCTGACGTGACGTTGTGTACTGTACTGT 1210
Db 18533 TTGTTGGGNGGTGTTTGGTNNNNGTTTNNNTTGTGNTGNGTGTGTGTTGTT 18592
Qy 1211 TGTGTTGTTTGTGTTG 1226
Db 18593 GTTGTGTTNTTTNTTG 18608

RESULT 18
AC130952/c
LOCUS
DEFINITION
AC130952
AC130952.1 GI:22261993
VERSION
KEYWORDS
SOURCE
ORGANISM
Rattus norvegicus
Rattus norvegicus
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
1 (bases 1 to 179590)
Munzy,D.Martie, Metker,M.Lee, Abramzon,S., Adams,C., Alder,J.,
Allen,C., Allen,H., Alebrooks,S., Amin,A., Angiano,D.,
Ayala-Bechl V., Aoyagi,A., Ayodeji,M., Bacc,E., Baden,H.,
Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,F.,
Biswalto,K., Blair,D., Blankenburg,K., Blyth,P., Brown,M.,
Bryant,N., Buha,C., Burch,P., Butrell,K., Calderon,E.,
Cardenas,V., Carter,K., Cavazos,I., Cesari,H., Center,A.,
Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Y., Chen,Z., Chu,J.,
Cleveland,L., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,L.,
Davis,M.L., Davis,C., Davy-Carroll,L., De Anda,C., Dederich,D.,
Delgado,O., Denson,S., Deramo,C., Ding,Y., Dinh,H., Divya,K.,
Draper,H., Dugan-Rocha,S., Dunn,A., Durbin,K., Duval,B., Eaves,K.,
Egan,A., Escotto,M., Eugene,C., Evans,C.A., Falls,T., Fan,G.,
Fernandez,S., Finley,M., Flagg,N., Forbes,L., Foster,M., Foster,P.,
Fraser,C.M., Gabisi,A., Ganta,R., Garcia,A., Garner,T., Garza,M.,
Gharategeogis,E., Geer,K., Gill,R., Grady,M., Guerra,W., Guevara,W.,
Gunnarsson,P., Haaland,W., Hamill,C., Hamilton,C., Hamilton,K.,
Harvey,Y., Havlak,P., Hawes,A., Henderson,N., Hernandez,J.,
Hernandez,R., Hines,S., Hladu,S.L., Hodgson,A., Hughes,M.,
Hollins,B., Howells,S., Hulik,S., Hume,J., Idlebird,D., Jackson,A.,
Jackson,L., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Jolivet,A.,
Karpachy,S., Kelly,S., Kelly,S., Khan,Z., King,L., Kovac,C.,
Kowls,C., Kraft,C.L., Lebow,H., Levan,J., Lewis,L., Li,Z., Liu,J.,
Liu,J., Liu,W., Liu,Y., London,P., Longacre,S., Lopez,J.,
Lorensuhewa,L., Louisedge,H., Lozada,R.J., Lu,X., Ma,J.,
Maheshwari,M., Mahindartine,M., Mahmood,M., Malloy,K., Mangum,A.,
Mangum,B., Mapus,P., Martin,K., Martin,R., Martinez,E.,
Mawhiney,S., McLeod,M., McNeill,T., Meenen,E., Mlloevilevic,A.,
Miner,G., Mirza,M., Moore,S., Moore,S., Morgan,M., Morris,K.,
Morris,S., Mundasa,M., Murphy,M., Nair,L., Nankervis,C., Neal,D.,
Newton,N., Nguyen,N., Norris,S., Nwokilemeh,O., Okunou,G.,
Olarpunsaagoon,A., Pal,S., Parks,K., Pasternak,S., Paul,H.,
Perez,A., Perez,L., Pfankoch,C., Plopper,F., Polinder,A.,
Popovic,D., Primus,E., Pu,L., Puazo,M., Quiroz,J., Rechin,E.,
Reeves,K., Regier,M.A., Reigh,R., Reilly,B., Reilly,M., Ren,Y.,
Reuter,M., Richards,S., Riggs,F., Rives,C., Rodkey,T., Rojas,A.,
Rose,M., Rose,R., Ruiz,S.J., Sanders,W., Savery,G., Scherer,S.,
Scott,G., Shatman,S., Shen,H., Shetty,J., Shvartsbeyn,A.,
Slason,I., Sitter,C.D., Smajs,D., Speed,A., Sodergren,E.,
Song,X.-Z., Sorelle,R., Sosa,J., Steidle,M., Strong,R., Sutton,A.,
Svatek,A., Taboc,P., Taylor,C., Taylor,T., Thomas,N., Thomas,S.,
Tingey,A., Trejos,Z., Usmani,K., Valas,R., Vera,V., Villasea,D.,
Waldron,L., Walker,B., Wang,J., Wang,Q., Wang,S., Warren,J.,
Warren,R., Wei,X., White,P., Williams,G., Willson,R., Wliscyk,R.,

```

Wooden, H., Worley, K., Wright, D., Wright, R., Wu, J., Yakub, S., Yen, J., Yoon, L., Yoon, V., Yu, F., Zhang, J., Zhou, J., Zhou, X., Zhao, S., Dunn, D., von Niederhausern, A., Weiss, R., Smith, D.R., Holt, R.A., Smith, H.O., Weinstein, G. and Gibbs, R.A.

TITLE  
JOURNAL  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL

Unpublished  
2 (bases 1 to 179590)  
Rat Genome Sequencing Consortium.  
Direct Submission  
Submitted (15-AUG-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA

-----  
Genome Center  
Center: Baylor College of Medicine  
Center code: BCM  
Web site: <http://www.hgsc.bcm.tmc.edu/>  
Contact: [hgsc-help@bcm.tmc.edu](mailto:hgsc-help@bcm.tmc.edu)  
-----  
Project Information  
Center project name: KCCA  
Center clone name: CH230-3806  
-----  
Summary Statistics  
Sequencing vector: Plasmid;  
Chemistry: Dye-terminator Big Dye; 100% of reads  
Assembly program: Phrap; version 0.990329  
Consensus quality: 115062 bases at least Q40  
Consensus quality: 12252 bases at least Q30  
Consensus quality: 128195 bases at least Q20  
-----

\* NOTE: Estimated insert size may differ from sequence length  
\* (see [http://www.hgsc.bcm.tmc.edu/docs/genbank\\_drafc\\_data.html](http://www.hgsc.bcm.tmc.edu/docs/genbank_drafc_data.html)).  
\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 68 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved.

\* 1 1262: contig of 1262 bp in length  
\* 1263 1362: gap of unknown length  
\* 1363 2382: contig of 1020 bp in length  
\* 2383 2482: gap of unknown length  
\* 2483 3528: contig of 1046 bp in length  
\* 3529 3629: gap of unknown length  
\* 3629 5049: contig of 1421 bp in length  
\* 5050 5149: gap of unknown length  
\* 5150 6413: contig of 1264 bp in length  
\* 6414 6513: gap of unknown length  
\* 6514 7596: contig of 1083 bp in length  
\* 7597 7697: gap of unknown length  
\* 7697 9104: contig of 1408 bp in length  
\* 9105 9204: gap of unknown length  
\* 9205 10311: contig of 1107 bp in length  
\* 10312 10411: gap of unknown length  
\* 10412 11512: contig of 1101 bp in length  
\* 11513 11612: gap of unknown length  
\* 11613 12754: contig of 1142 bp in length  
\* 12755 12854: gap of unknown length  
\* 12855 13908: contig of 1054 bp in length  
\* 13909 14008: gap of unknown length  
\* 14009 15562: contig of 1554 bp in length  
\* 15563 15662: gap of unknown length  
\* 15663 17214: contig of 1552 bp in length  
\* 17215 17314: gap of unknown length  
\* 17315 18574: contig of 1260 bp in length  
\* 18575 18674: gap of unknown length  
\* 18675 19719: contig of 1045 bp in length  
\* 19720 19819: gap of unknown length  
\* 19820 21191: contig of 1372 bp in length  
\* 21192 21291: gap of unknown length  
\* 21292 22376: contig of 1085 bp in length  
\* 22377 22476: gap of unknown length  
\* 22477 23555: contig of 1079 bp in length

\* 23556 23656: gap of unknown length  
\* 23656 25139: contig of 1484 bp in length  
\* 25140 25239: gap of unknown length  
\* 25240 27004: contig of 1765 bp in length  
\* 27005 27104: gap of unknown length  
\* 27105 29050: contig of 1946 bp in length  
\* 29051 30697: gap of unknown length  
\* 30697 30796: contig of 1546 bp in length  
\* 30797 31831: gap of unknown length  
\* 31831 31931: contig of 1034 bp in length  
\* 31931 34241: gap of unknown length  
\* 34242 34341: contig of 211 bp in length  
\* 34341 35956: gap of unknown length  
\* 35957 36056: contig of 1615 bp in length  
\* 36057 37586: gap of unknown length  
\* 37587 37686: contig of 1530 bp in length  
\* 37687 39221: gap of unknown length  
\* 39222 39321: contig of 1535 bp in length  
\* 39322 40412: gap of unknown length  
\* 40413 40512: contig of 1091 bp in length  
\* 40513 41662: gap of unknown length  
\* 41663 41762: contig of 1150 bp in length  
\* 41763 43490: gap of unknown length  
\* 43491 43590: contig of 1728 bp in length  
\* 43591 45380: gap of unknown length  
\* 45381 45480: contig of 1790 bp in length  
\* 45481 47099: gap of unknown length  
\* 47099 47100: contig of 1619 bp in length  
\* 47100 47200: gap of unknown length  
\* 47200 49229: contig of 2029 bp in length  
\* 49229 49329: gap of unknown length  
\* 49329 50611: contig of 1282 bp in length  
\* 50611 50710: contig of 1441 bp in length  
\* 50711 52152: gap of unknown length  
\* 52152 52251: contig of 141 bp in length  
\* 52251 53269: gap of unknown length  
\* 53269 53370: contig of 1018 bp in length  
\* 53370 55118: gap of unknown length  
\* 55118 55619: contig of 2149 bp in length  
\* 55619 55619: gap of unknown length  
\* 55619 58162: contig of 2544 bp in length  
\* 58163 58262: gap of unknown length  
\* 58262 61562: gap of unknown length  
\* 61563 61662: contig of 3300 bp in length  
\* 61663 63096: gap of unknown length  
\* 63097 63196: contig of 1434 bp in length  
\* 63197 65617: gap of unknown length  
\* 65618 65717: contig of 2421 bp in length  
\* 65718 68569: gap of unknown length  
\* 68570 68669: contig of 2852 bp in length  
\* 68670 70858: gap of unknown length  
\* 70859 70958: contig of 2189 bp in length  
\* 70959 73184: gap of unknown length  
\* 73185 73284: contig of 2226 bp in length  
\* 73285 76201: gap of unknown length  
\* 76202 76301: contig of 2917 bp in length  
\* 76302 79387: gap of unknown length  
\* 79388 79487: contig of 3086 bp in length  
\* 79488 81544: gap of unknown length  
\* 81545 81644: contig of 2057 bp in length  
\* 81645 84934: gap of unknown length  
\* 84935 85034: contig of 3290 bp in length  
\* 85035 88741: gap of unknown length  
\* 88742 88842: contig of 3707 bp in length  
\* 88843 92553: gap of unknown length  
\* 92554 92653: contig of 3712 bp in length  
\* 92654 96858: gap of unknown length  
\* 96859 96959: contig of 4205 bp in length  
\* 96960 100191: gap of unknown length  
\* 100192 100291: contig of 3233 bp in length  
\* 100292 103270: gap of unknown length  
\* 103271 103370: contig of 2979 bp in length  
\* 103370: gap of unknown length

Query Match

3.9%; Score 57.8; DB 2; Length 179590;







Chen, G., Chen, R., Chen, Z., Chowdhry, I., Christopoulos, C., Cleveland, C.D., Cox, C., Coyle, M.D., Dathorne, S.R., David, R., Davila, M.L., Davis, C., Davy-Carroll, L., Dederich, D.A., Delaney, K.R., Delgado, O., Dem, A.L., Ding, Y., Dinh, H.H., Douthett, K.J., Draper, H., Dugan-Rocha, S., Durbin, K.J., Earnhart, C., Edgar, D., Edwards, C.C., Elhaj, C., Escotto, M., Falls, T., Ferraguto, D., Flagg, N., Ford, J., Foster, P., Frintz, P., Gabali, A., Gao, J., Garcia, A., Garner, T., Garza, N., Gill, R., Gorrell, J.H., Guevara, W., Gunaratne, P., Hale, S., Hamilton, K., Harris, C., Harris, K., Hart, M., Havlak, P., Hawes, A., Hernandez, J., Hernandez, O., Hodgson, A., Hogues, M., Holloway, C., Hollins, B., Homai, F., Howard, S., Huber, J., Huiyk, S., Hume, J., Jackson, L.E., Jacobson, B., Jia, Y., Johnson, R., Jolivet, S., Joudah, S., Karlsson, E., Kelly, S., Khan, U., Kling, L., Korvah, J., Kovar, C., Kratovic, J., Kureshi, A., Landry, N., Leal, B., Lewis, L.C., Lewis, L., Li, Z., Lichtarge, O., Lieu, C., Liu, J., Liu, W., Louised, H., Lorado, R.J., Lu, X., Lucier, A., Lucier, R., Luna, R., Ma, J., Maheshwari, M., Mapua, P., Martin, R., Martindale, A., Martinez, E., Massey, E., Mawhiney, E., McLeod, M.P., Meador, M., Mei, G., Metzker, M., Miner, G., Miner, Z., Mitchell, T., Mohabbat, K., Morgan, M., Morris, S., Moser, M., Neal, D., Newton, J., Newton, N., Nguyen, A., Nguyen, N., Nguyen, N., Nickerson, E., Nwokenkwo, S., Ogih, M., Okunolu, G., Oragune, N., Oviedo, R., Pace, A., Payton, B., Peery, J., Perez, L., Peters, L., Pickens, R., Primus, E., Pu, L.L., Quiles, M., Ren, Y., Rives, M., Rojas, A., Rojubo, I., Rolfe, M., Ruiz, S., Savery, G., Scherer, S., Scott, G., Shen, H., Shochat, N., Sisson, I., Sodergren, E., Sonalke, T., Sparks, A., Stanley, H., Stone, H., Sutton, A., Svatek, A., Taber, P., Tamerisa, A., Tamerisa, K., Tang, H., Tansey, J., Taylor, C., Taylor, T., Telford, B., Thomas, N., Thomas, S., Usmani, K., Vaequez, L., Vera, V., Villalón, D., Vinson, R., Wang, Q., Wang, S., Ward-Moore, S., Warren, R., Washington, C., Watlington, S., Williams, G., Williamson, A., Wleczek, R., Wooden, S., Worley, K., Wu, C., Wu, Y., Wu, Y.F., Zhou, J., Zorrilla, S., Nelson, D., Weinstein, G. and Gibbs, R.

Direct Submission  
Unpublished  
2 (bases 1 to 132168)  
Worley, K.C.

Direct Submission  
Submitted (26-Apr-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA  
3 (bases 1 to 132168)  
Worley, K.C.

Direct Submission  
Submitted (18-Jul-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA  
On Jul 14, 2002 this sequence version replaced gi:20336688.

----- Genome Center  
Center: Baylor College of Medicine  
Center code: BCM  
Web site: <http://www.hgsc.bcm.tmc.edu/>  
Contact: hgsc-help@bcm.tmc.edu  
----- Project Information  
Center project name: GUMU  
Center clone name: CH230-242E2  
----- Summary Statistics  
Sequencing vector: Plasmid  
Chemistry: Dye-terminator Big Dye 100% of reads  
Assembly program: Phrap, version 0.990329  
Consensus quality: 65480 bases at least Q40  
Consensus quality: 68839 bases at least Q30  
Consensus quality: 71356 bases at least Q20

-----  
\* NOTE: Estimated insert size may differ from sequence length  
\* (see <http://www.hgsc.bcm.tmc.edu/docs/genbankdraftdata.html>).  
\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 59 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence

\* as soon as it is available and the accession number will  
\* be preserved.

1	1059:	contig of 1059 bp in length
*	1060:	gap of unknown length
*	1159:	contig of 1086 bp in length
*	1245:	contig of 1086 bp in length
*	2245:	gap of unknown length
*	2345:	gap of unknown length
*	3365:	contig of 1020 bp in length
*	3366:	gap of unknown length
*	3367:	gap of unknown length
*	4459:	contig of 1084 bp in length
*	4460:	gap of unknown length
*	4650:	contig of 1866 bp in length
*	6535:	gap of unknown length
*	6536:	gap of unknown length
*	6636:	contig of 1100 bp in length
*	7735:	contig of 1100 bp in length
*	7835:	gap of unknown length
*	9146:	contig of 1310 bp in length
*	9245:	gap of unknown length
*	9246:	contig of 1318 bp in length
*	10563:	gap of unknown length
*	10663:	gap of unknown length
*	12288:	contig of 1625 bp in length
*	12388:	gap of unknown length
*	13840:	contig of 1452 bp in length
*	13940:	gap of unknown length
*	13941:	contig of 1039 bp in length
*	14979:	contig of 1039 bp in length
*	15079:	gap of unknown length
*	15080:	contig of 1322 bp in length
*	15081:	contig of 1322 bp in length
*	16402:	gap of unknown length
*	16501:	gap of unknown length
*	16502:	contig of 1566 bp in length
*	18067:	contig of 1566 bp in length
*	18167:	gap of unknown length
*	18168:	contig of 1846 bp in length
*	20013:	contig of 1846 bp in length
*	20113:	gap of unknown length
*	21219:	contig of 1106 bp in length
*	21319:	gap of unknown length
*	22874:	contig of 1555 bp in length
*	22974:	gap of unknown length
*	22975:	gap of unknown length
*	24072:	contig of 1098 bp in length
*	24073:	gap of unknown length
*	24172:	gap of unknown length
*	25306:	contig of 1134 bp in length
*	25406:	gap of unknown length
*	25407:	gap of unknown length
*	25790:	contig of 1384 bp in length
*	26791:	gap of unknown length
*	26891:	contig of 1373 bp in length
*	28263:	gap of unknown length
*	28264:	gap of unknown length
*	29935:	contig of 1572 bp in length
*	29936:	gap of unknown length
*	30035:	gap of unknown length
*	31819:	contig of 1784 bp in length
*	31919:	gap of unknown length
*	31920:	gap of unknown length
*	33826:	contig of 1907 bp in length
*	33827:	gap of unknown length
*	33926:	gap of unknown length
*	35427:	contig of 1501 bp in length
*	35428:	gap of unknown length
*	35527:	gap of unknown length
*	35529:	contig of 1032 bp in length
*	36559:	gap of unknown length
*	36560:	gap of unknown length
*	36660:	contig of 1742 bp in length
*	38401:	gap of unknown length
*	38402:	gap of unknown length
*	38502:	contig of 1090 bp in length
*	39591:	gap of unknown length
*	39592:	gap of unknown length
*	40985:	contig of 1295 bp in length
*	41086:	gap of unknown length
*	41087:	gap of unknown length
*	42622:	contig of 1536 bp in length
*	42722:	gap of unknown length
*	44518:	contig of 1796 bp in length
*	44618:	gap of unknown length
*	44619:	gap of unknown length
*	4560:	contig of 1942 bp in length
*	46660:	gap of unknown length
*	46661:	gap of unknown length
*	48640:	contig of 1980 bp in length
*	48641:	gap of unknown length
*	48741:	contig of 1823 bp in length
*	50563:	contig of 1823 bp in length
*	50663:	gap of unknown length
*	52763:	contig of 2000 bp in length
*	52764:	gap of unknown length
*	52765:	gap of unknown length
*	54038:	contig of 1275 bp in length
*	54138:	gap of unknown length
*	56145:	contig of 2007 bp in length





[illegible]

RESULT 22					
LOCUS	AC128718				
DEFINITION	AC128718	172387 bp	DNA	linear	HTG 02-AUG-2002
ACCESSION	Rattus norvegicus clone CH230-141N16,				
VERSION	***, 44 unordered pieces.				
KEYWORDS	AC128718				
SOURCE	AC128718.2	GI:22038351			
ORGANISM	HTG, HTGS, PHASE1.				
	Norway rat.				
	Rattus norvegicus				
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;				
	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;				
	Rattus				
REFERENCE	1 (bases 1 to 172387)				
AUTHORS	Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C.,				
	Alsbrooks,S.L., Amarantunge,H.C., Are,J.R., Ayale,M., Banks,T.,				
	Barbarta,J., Benton,J., Binage,K., Blankenburg,K., Bonini,D.,				
	Bouck,J., Bowls,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P.,				
	Buhay,C., Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C.,				
	Carroll,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D.,				
	Chen,G., Chen,R., Chen,Z., Chowdhry,I., Christopoulos,C.,				
	Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,D.A., David,R.,				
	Devila,M.L., Davis,C., Davy-Carroll,L., Dederich,D.A.,				
	Delaney,K.R., Delgado,O., Demn,A.L., Ding,Y., Dinl,H.H.,				
	Douthwaite,K.J., Draper,H., Dugan-Rocha,S., Dublin,K.J.,				
	Earnhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Escotto,M.,				
	Falls,T., Ferraguto,D., Flagg,N., Ford,U., Foster,P., Frantz,P.,				
	Gabisi,T., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R.,				
	Gorelli,J.H., Guevara,M., Gunaratne,P., Hale,S., Hamilton,K.,				
	Harris,C., Harris,K., Hart,M., Haylak,P., Hawes,A., Hernandez,J.,				
	Hernandez,O., Hodgson,A., Hognes,M., Holloway,C., Hollins,B.,				
	Homesi,F., Howard,S., Huber,J., Hultyk,S., Hune,U., Jackson,L.E.,				
	Jacobson,B., Jia,Y., Johnson,R., Jolivet,S., Joudah,S.,				
	Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J., Kovar,C.,				
	Kracovic,J., Kureishi,A., Landry,N., Leal,B., Lewis,L.C., Lewis,L.,				
	Li,J., Li,Z., Lichtarge,O., Lien,C., Liu,J., Liu W., Louisged,H.,				
	Lozdo,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R., Ma,J.,				
	Maheshwari,M., Mapua,P., Martin,R., Martindale,A., Martinez,E.,				
	Maseley,B., Mawliny,E., Mcleod,M.P., Meador,M., Mei,G., Metzger,M.,				

TITLE	
JOURNAL	
REFERENCE	
AUTHORS	
TITLE	
JOURNAL	
REFERENCES	
AUTHORS	
TITLE	
JOURNAL	
COMMENT	

Miner, G., Inner, Z., Mitchell, T., Mohabbat, K., Morgan, M., Morris, S., Moser, M., Neal, D., Newton, J., Newton, N., Nguyen, A., Nguyen, N., Nguyen, N., Nickerson, E., Nwokweno, S., Ogun, M., Okunribido, A., Oragunye, N., Oyedero, R., Pace, A., Payton, B., Peery, J., Perez, L., Peters, L., Pickens, R., Primus, E., Pu, L., Qureshi, M., Ren, Y., Rivers, M., Rojas, A., Rojibokan, I., Rolfe, M., Ruiz, S., Savery, G., Scheer, S., Scott, G., Shen, H., Shoohtari, N., Sisson, I., Sodeghien, E., Sonake, T., Sparks, A., Stanley, H., Stone, H., Sutton, A., Svatek, A., Tabors, P., Tamerisa, A., Tamerisa, K., Tang, H., Tansey, J., Taylor, C., Taylor, T., Teifrod, B., Thomas, N., Thomas, S., Usmani, K., Vasquez, L., Vera, V., Villalón, D., Vinson, R., Wang, Q., Wang, S., Ward-Moore, S., Warren, R., Washington, C., Watlington, S., Williams, G., Williamson, A., Wleciyk, R., Wooden, S., Worley, K., Wu, C., Wu, Y., Wu, Y., Zhou, D., Zorrilla, S., Nelson, D., Weinstein, G., and Gibbs, R.

Direct Submission  
Unpublished  
2 (bases 1 to 172387)  
Worley, K.C.  
Direct Submission  
Submitted (23-JUL-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA  
3 (bases 1 to 172387)  
Worley, K.C.  
Direct Submission  
Submitted (02-AUG-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA  
On Aug 1, 2002 this sequence version replaced gi:21930181.

```

Center clone name: CH330-141N16
----- Summary Statistics -----
Sequencing Vector: Plasmid;
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.990329
Consensus quality: 133362 bases at least Q40
Consensus quality: 139249 bases at least Q30
Consensus quality: 144009 bases at least Q20

-----
* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/genbank\_drat1\_data.html)
* NOTE: This is a 'working draft' sequence. It currently
* consists of 44 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

1
1157 1156: contig of 1156 bp in length
1257 1256: gap of unknown length
1257 2433: contig of 1177 bp in length
2434 2533: gap of unknown length
2534 2534 3617: contig of 1084 bp in length
3618 3717: gap of unknown length
3718 5064: contig of 1347 bp in length
5065 5166: gap of unknown length
5166 6436: contig of 1272 bp in length
6437 6536: gap of unknown length
6537 8134: contig of 1598 bp in length
8135 8235 9794: contig of 1560 bp in length
9795 9895: gap of unknown length
9896 11641: contig of 1747 bp in length
11642 11642 13093: contig of 1352 bp in length

```



Gabisli, A., Gao, J., Garcia, A., Garner, T., Garza, N., Gill, R.,  
Gorrell, J. H., Guevara, W., Gunaratne, P., Hale, S., Hamilton, K.,  
Harris, C., Harris, K., Hart, M., Havlak, P., Hawes, A., Hernandez, J.,  
Henningson, O., Hodgson, A., Hogues, M., Holloway, C., Hollins, B.,  
Honsi, F., Howard, S., Huber, J., Huliy, S., Hume, J., Jackson, L. E.,  
Jacobson, B., Jia, Y., Johnson, R., Jolivet, S., Joudah, S.,  
Karlsom, E., Kelly, S., Khan, U., King, L., Korvah, J., Kovar, C.,  
Kratovic, J., Kureshi, A., Landry, N., Leal, B., Lewis, L. C., Lewis, L.,  
Li, J., Li, Z., Lichtarge, O., Lieu, C., Liu, J., Liu, W., Louised, H.,  
Lorado, R. J., Lu, X., Lucier, A., Lucier, R., Luna, R., Ma, J.,  
Maneshwari, M., Mapa, P., Martin, R., Martindale, A., Martinez, E.,  
Massey, E., Mawhinney, E., McLeod, M. P., Meador, M., Mei, G., Metcalf, M.,  
Miner, G., Miner, Z., Mitchell, T., Mohabbat, K., Morgan, M., Morris, S.,  
Moser, N., Neal, D., Newton, J., Newton, N., Nguyen, A., Nguyen, N.,  
Nguyen, N., Nickerson, E., Nwokenkwo, S., Oguh, M., Okunolu, G.,  
Orgunye, N., Oviedo, R., Pace, A., Payton, B., Peary, J., Perez, L.,  
Peters, L., Pickens, R., Primus, E., Pu, L. L., Quiles, M., Ren, Y.,  
Rivers, M., Rojas, A., Rojoudokan, I., Rolfe, M., Ruiz, S., Savery, G.,  
Scherer, S., Scott, G., Shen, H., Shooshart, N., Sisson, I.,  
Sodergren, E., Sonaike, T., Sparks, A., Stanley, H., Stone, H.,  
Sutton, A., Swatek, A., Taber, P., Tamerisa, A., Tamerisa, K., Tang, H.,  
Tansey, J., Taylor, C., Taylor, T., Telford, B., Thomas, N., Thomas, S.,  
Usmani, K., Vasquez, L., Vera, V., Villalon, D., Vinson, R., Wang, Q.,  
Wang, S., Ward-Moore, S., Warren, R., Washington, C., Watlington, S.,  
Williams, G., Williamson, A., Wlarczyk, R., Wooden, S., Worley, K.,  
Wu, C., Wu, Y., Wu, Y. F., Zhou, J., Zorrilla, S., Nelson, D.,  
Weinstock, G. and Gibbs, R.

Unpublished  
Direct Submission  
2 (bases 1 to 159876)  
Worley, K. C.

Direct Submission  
Submitted (05-MAR-2002) Human Genome Sequencing Center, Department  
of Molecular and Human Genetics, Baylor College of Medicine, One  
Baylor Plaza, Houston, TX 77030, USA  
3 (bases 1 to 159876)  
Worley, K. C.

Direct Submission  
Submitted (17-JUN-2002) Human Genome Sequencing Center, Department  
of Molecular and Human Genetics, Baylor College of Medicine, One  
Baylor Plaza, Houston, TX 77030, USA  
On Jul 13, 2002 this sequence version replaced gi:19525875.

Genome Center  
Center: Baylor College of Medicine  
Center code: BCM  
Web site: <http://www.hgsc.bcm.tmc.edu/>  
Contact: hgsc-help@bcm.tmc.edu  
Project Information  
Center project name: GSBW  
Center clone name: CH230-57B10  
Summary Statistics  
Sequencing vector: plasmid:  
Chemistry: Dye-terminator Big Dye 100% of reads  
Assembly program: Phrap; version 0.990329  
Consensus quality: 121438 bases at least Q40  
Consensus quality: 127117 bases at least Q30  
Consensus quality: 131313 bases at least Q20

NOTE: Estimated insert size may differ from sequence length  
(see [http://www.hgsc.bcm.tmc.edu/docs/genbank\\_draht\\_data.html](http://www.hgsc.bcm.tmc.edu/docs/genbank_draht_data.html)).  
NOTE: This is a "working draft" sequence. It currently  
consists of 63 contigs. The true order of the pieces  
is not known and their order in this sequence record is  
arbitrary. Gaps between the contigs are represented as  
runs of N, but the exact sizes of the gaps are unknown.  
This record will be updated with the finished sequence  
as soon as it is available and the accession number will  
be preserved.

1 1123: contig of 1123 bp in length  
\* 1124 2306: gap of unknown length  
\* 1224 2306: contig of 1083 bp in length  
\* 2307 2406: gap of unknown length  
\* 2407 3650: contig of 1244 bp in length

3651 3750: gap of unknown length  
\* 3751 5094: contig of 1344 bp in length  
\* 5095 5194: gap of unknown length  
\* 5195 6987: contig of 1793 bp in length  
\* 6988 7088: gap of unknown length  
\* 7089 8124: contig of 1037 bp in length  
\* 8125 8225: gap of unknown length  
\* 8226 9974: contig of 1750 bp in length  
\* 9975 10074: gap of unknown length  
\* 10075 11134: contig of 1060 bp in length  
\* 11135 11234: gap of unknown length  
\* 11235 13374: contig of 2140 bp in length  
\* 13375 14744: gap of unknown length  
\* 14745 14522: contig of 1048 bp in length  
\* 14523 14622: gap of unknown length  
\* 14623 15896: contig of 1274 bp in length  
\* 15897 15996: gap of unknown length  
\* 15997 17518: contig of 1522 bp in length  
\* 17519 17618: gap of unknown length  
\* 17619 16970: contig of 1352 bp in length  
\* 16971 19070: gap of unknown length  
\* 19071 21625: contig of 2555 bp in length  
\* 21626 21725: gap of unknown length  
\* 21726 22854: contig of 1129 bp in length  
\* 22855 22954: gap of unknown length  
\* 22955 24344: contig of 1390 bp in length  
\* 24345 24444: gap of unknown length  
\* 24445 26228: contig of 1784 bp in length  
\* 26229 26328: gap of unknown length  
\* 26329 26774: contig of 2346 bp in length  
\* 26775 28774: gap of unknown length  
\* 28775 30141: contig of 1367 bp in length  
\* 30142 30241: gap of unknown length  
\* 30242 32024: contig of 1783 bp in length  
\* 32025 32124: gap of unknown length  
\* 32125 33926: contig of 1802 bp in length  
\* 33927 34026: gap of unknown length  
\* 34027 36237: contig of 2211 bp in length  
\* 36238 36337: gap of unknown length  
\* 36339 37818: contig of 1481 bp in length  
\* 37819 37919: gap of unknown length  
\* 37919 38825: contig of 1007 bp in length  
\* 38825 39025: gap of unknown length  
\* 39026 40937: contig of 1912 bp in length  
\* 40938 41037: gap of unknown length  
\* 41038 42502: contig of 1465 bp in length  
\* 42503 42602: gap of unknown length  
\* 42603 44552: contig of 1850 bp in length  
\* 44553 44552: gap of unknown length  
\* 44553 46246: contig of 1694 bp in length  
\* 46247 46346: gap of unknown length  
\* 46347 49260: contig of 2914 bp in length  
\* 49261 49360: gap of unknown length  
\* 49361 51694: contig of 2334 bp in length  
\* 51695 51794: gap of unknown length  
\* 51795 53387: contig of 1493 bp in length  
\* 53388 55316: gap of unknown length  
\* 55317 55416: contig of 1929 bp in length  
\* 55417 57061: gap of unknown length  
\* 57062 57161: gap of unknown length  
\* 57162 59882: contig of 2721 bp in length  
\* 59883 59982: gap of unknown length  
\* 59983 61398: contig of 1416 bp in length  
\* 61399 61496: gap of unknown length  
\* 61496 64096: contig of 2588 bp in length  
\* 64097 64196: gap of unknown length  
\* 64197 65752: contig of 1556 bp in length  
\* 65753 65852: gap of unknown length  
\* 65853 68376: contig of 2424 bp in length  
\* 68377 71358: gap of unknown length  
\* 71359 71458: gap of unknown length







ACCESSION AC098913  
 VERSION AC098913.4 GI:21730008  
 KEYWORDS HTG; HTGS PHASE1.  
 SOURCE Norway rat.  
 ORGANISM Rattus norvegicus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;  
 Rattus.

REFERENCE 1 (bases 1 to 168382)  
 AUTHORS Murthy D.M., Adams C., Adio-Oduola B., Ali-osman, F.R., Allen C.,  
 Albrooks S.L., Amaralunge H.C., Are J.R., Ayalew M., Banks T.,  
 Barbarta J., Benton J., Blinage K., Blankenburg K., Bonini D.,  
 Bouck J., Bowle S., Brileva M., Brown E., Brown M., Bryant N.P.,  
 Buhay C., Burch P., Burkett C., Burrell K.L., Byrd N.C.,  
 Carron T.F., Carter M., Cavazos S.R., Chacko J., Chavez D.,  
 Chen G., Chen R., Chen Z., Chowdhry I., Christopoulos C.,  
 Cleveland C.D., Cox C., Coyle M.D., Dathorne S.R., David R.,  
 Davila M.L., Davis C., Davy-Carroll L., Dederich D.A.,  
 Delaney K.R., Delgado O., Denn A.L., Ding Y., Dinh H.H.,  
 Douthwaite K.J., Draper H., Dugan-Rocha S., Durbin K.J.,  
 Earhart C., Edgar D., Edwards C.C., Elhaj C., Escotto M.,  
 Falls T., Ferraguto D., Flagg N., Ford J., Foster P., Frantz P.,  
 Gabisi A., Gao U., Garcia A., Garner T., Garza N., Gill R.,  
 Gorrell J.H., Guevara W., Gunaratne P., Hale S., Hamilton K.,  
 Harris C., Harris K., Hart M., Havlak P., Hawes A., Hernandez J.,  
 Hernandez O., Hodgson A., Hognes M., Holloway C., Hollins B.,  
 Homai P., Howard S., Huber J., Hulik S., Hume J., Jackson L.E.,  
 Jacobsen B., Jia Y., Johnson R., Jolivet S., Joudah S.,  
 Karlsson E., Kelly S., Khan U., King L., Kovach J., Kovar C.,  
 Krtolovic J., Kureshi A., Landry N., Leal B., Lewis L.C., Lewis L.,  
 Li J., Li Z., Lichtarge O., Lieu C., Liu J., Liu W., Louised H.,  
 Llado R.J., Lu X., Lucier A., Lucier R., Luna R., Ma J.,  
 Mareshwar M., Mapa P., Martin R., Martindale A., Matinez E.,  
 Massey E., McWhiney E., McLeod M.P., Meador M., Mei G., Metcalf M.,  
 Miner G., Miner Z., Mitchell T., Mohabbat K., Morgan M., Morris S.,  
 Moser M., Neal D., Newton J., Newton N., Nguyen A., Nguyen N.,  
 Nguyen N., Nickerson E., Nwokenwu S., Ogih M., Okunolu G.,  
 Orangun N., Oviedo R., Pace A., Payton B., Peery J., Perez L.,  
 Peters L., Pickens R., Primus E., Pu L., Quiles M., Ren Y.,  
 Rives M., Rojas A., Rojokokan I., Rolfe M., Ruiz S., Savary G.,  
 Scherter S., Scott G., Shen H., Shoohtari N., Sisson I.,  
 Sodergren E., Sotaike T., Sparks A., Stanley H., Stone H.,  
 Sutton A., Syatek A., Tabor P., Tamerisa A., Tamerisa K., Tang H.,  
 Tansey J., Taylor C., Taylor T., Teleford B., Thomas N., Thomas S.,  
 Umani K., Vasquez L., Vera V., Villalon D., Vinson R., Wang Q.,  
 Wang S., Ward-Moore S., Warren R., Washington C., Watling K.,  
 Williams G., Williamson A., Wleczek R., Wooden S., Worley K.,  
 Wu C., Wu Y., Wu Y.F., Zhou J., Zorrilla S., Nelson D.,  
 Weinstock G. and Glibbs R.

TITLE Direct Submission  
 JOURNAL Unpublished  
 REFERENCE 2 (bases 1 to 168382)  
 AUTHORS Worley K.C.  
 TITLE Direct Submission  
 JOURNAL Submitted (05-NOV-2001) Human Genome Sequencing Center, Department  
 of Molecular and Human Genetics, Baylor College of Medicine, One  
 Baylor Plaza, Houston, TX 77030, USA  
 3 (bases 1 to 168382)  
 AUTHORS Worley K.C.  
 TITLE Direct Submission  
 JOURNAL Submitted (12-JUL-2002) Human Genome Sequencing Center, Department  
 of Molecular and Human Genetics, Baylor College of Medicine, One  
 Baylor Plaza, Houston, TX 77030, USA  
 On Jul 11, 2002 this sequence version replaced gi:17974091.

COMMENT ----- Genome Center  
 Center: Baylor College of Medicine  
 Center code: BCM  
 Web site: <http://www.hgsc.bcm.tmc.edu/>  
 Contact: hgsc-help@bcm.tmc.edu  
 ----- Project Information  
 Center Project name: GIGS  
 Center clone name: CH230-182P23  
 ----- Summary Statistics

Sequencing vector: Plasmid;  
 Chemistry: Dye-terminator Big Dye; 100% of reads  
 Assembly program: Phrap; version 0.990329  
 Consensus quality: 102338 bases at least Q40  
 Consensus quality: 108995 bases at least Q30  
 Consensus quality: 113860 bases at least Q20

-----  
 \* NOTE: Estimated insert size may differ from sequence length  
 \* (see [http://www.hgsc.bcm.tmc.edu/docs/genbank\\_draft\\_data.html](http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html)).  
 \* NOTE: This is a "working draft" sequence. It currently  
 \* consists of 80 contigs. The true order of the pieces  
 \* is not known and their order in this sequence record is  
 \* arbitrary. Gaps between the contigs are represented as  
 \* runs of N, but the exact sizes of the gaps are unknown.  
 \* This record will be updated with the finished sequence  
 \* as soon as it is available and the accession number will  
 \* be preserved.

1	1161:	contig of 1161 bp in length
1162	1261:	gap of unknown length
1262	1458:	contig of 1197 bp in length
1263	2458:	gap of unknown length
1264	2558:	gap of unknown length
1265	3699:	contig of 1141 bp in length
1266	3799:	gap of unknown length
1267	3800:	contig of 1034 bp in length
1268	4834:	gap of unknown length
1269	4934:	contig of 1173 bp in length
1270	6106:	gap of unknown length
1271	6206:	gap of unknown length
1272	7621:	contig of 1415 bp in length
1273	7721:	gap of unknown length
1274	8807:	contig of 1086 bp in length
1275	8907:	gap of unknown length
1276	10076:	contig of 1169 bp in length
1277	10176:	gap of unknown length
1278	11821:	contig of 1645 bp in length
1279	11921:	gap of unknown length
1280	13381:	contig of 1460 bp in length
1281	13481:	gap of unknown length
1282	14669:	contig of 1188 bp in length
1283	14769:	gap of unknown length
1284	16044:	contig of 1275 bp in length
1285	16144:	gap of unknown length
1286	16145:	gap of unknown length
1287	17231:	contig of 1087 bp in length
1288	17331:	gap of unknown length
1289	18776:	contig of 1445 bp in length
1290	18876:	gap of unknown length
1291	20096:	contig of 1220 bp in length
1292	20196:	gap of unknown length
1293	21546:	contig of 1350 bp in length
1294	21646:	gap of unknown length
1295	23191:	contig of 1545 bp in length
1296	23291:	gap of unknown length
1297	24672:	contig of 1381 bp in length
1298	24772:	gap of unknown length
1299	26145:	contig of 1373 bp in length
1300	26245:	gap of unknown length
1301	27467:	contig of 1222 bp in length
1302	27567:	gap of unknown length
1303	28967:	contig of 1400 bp in length
1304	29067:	gap of unknown length
1305	30131:	contig of 1064 bp in length
1306	30231:	gap of unknown length
1307	31463:	contig of 1232 bp in length
1308	31563:	gap of unknown length
1309	32893:	contig of 1330 bp in length
1310	32993:	gap of unknown length
1311	34539:	contig of 1546 bp in length
1312	34639:	gap of unknown length
1313	36140:	contig of 1501 bp in length
1314	36240:	gap of unknown length
1315	36749:	contig of 2509 bp in length
1316	38849:	gap of unknown length
1317	39857:	contig of 1008 bp in length
1318	39957:	gap of unknown length

```
* 3958      41409: contig of 1452 bp in length
* 41510     41509: gap of unknown length
* 43123     43123: contig of 1614 bp in length
* 43224     43223: gap of unknown length
* 43324     43324: contig of 1168 bp in length
* 44392     44391: gap of unknown length
* 44492     44827: contig of 2336 bp in length
* 46828     46927: gap of unknown length
* 46928     48247: contig of 1320 bp in length
* 48248     48347: gap of unknown length
* 48348     49512: contig of 1165 bp in length
* 49513     49612: gap of unknown length
* 49613     51144: contig of 1532 bp in length
* 51145     51244: gap of unknown length
* 51245     52796: contig of 1552 bp in length
* 52797     52896: gap of unknown length
* 52897     54388: contig of 1492 bp in length
* 54389     54488: gap of unknown length
* 54489     55993: contig of 1505 bp in length
* 55994     56093: gap of unknown length
* 56094     58690: contig of 2597 bp in length
* 58691     58790: gap of unknown length
* 58791     60218: contig of 1428 bp in length
* 60219     60318: gap of unknown length
* 60319     61518: contig of 1200 bp in length
* 61519     61618: gap of unknown length
* 61619     63861: contig of 2243 bp in length
* 63862     63961: gap of unknown length
* 63962     66472: contig of 2511 bp in length
* 66473     66572: gap of unknown length
* 66573     68301: contig of 1729 bp in length
* 68302     68401: gap of unknown length
* 68402     69700: contig of 1299 bp in length
* 69701     69800: gap of unknown length
* 69801     71414: contig of 1614 bp in length
* 71415     71514: gap of unknown length
* 71515     73479: contig of 1965 bp in length
* 73480     73579: gap of unknown length
* 73580     75907: contig of 2328 bp in length
* 75908     76007: gap of unknown length
* 76008     78596: contig of 2589 bp in length
* 78597     78696: gap of unknown length
* 78697     80202: contig of 1506 bp in length
* 80203     80302: gap of unknown length
* 80303     81912: contig of 1610 bp in length
* 81913     82012: gap of unknown length
* 82013     83360: contig of 1348 bp in length
* 83361     83460: gap of unknown length
* 83461     85116: contig of 1656 bp in length
* 85117     85216: gap of unknown length
* 85217     87725: contig of 2509 bp in length
```

Query Match 3.8%; Score 56; DB 2; Length 168382;  
Best Local Similarity 46.5%; Pred. No. 8.1e-05;  
Matches 158; Conservative 0; Mismatches 182; Indels 0; Gaps 0;

```
QY 886 GAAATGAGTGAATGAGTGAATGAGACCACTGGAGAGAGAAAGCTACATGAATAATCT 945
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 14313 GACTCACTAGTGAGTAATATATAGTAAATTTAATTAATAAAAAAGTATTCCTCT 14372

QY 946 ATGCAAAACATGAGAAATGCTGCTGTACATGAAGTACGGGGCTCAATGCTTTGAGTATGT 1005
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 14373 ACCAATTTGCTTCTCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 14432

QY 1006 GTTTGGAGAGTGTGAGATGAGATGACTTGAATCGTCAATATATGACATGATGATTG 1065
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 14433 NTGNGTNTTTGTGTGNGGNGGTGTGTTGTTTGTGTTGGGGGTGTTNNNG 14492

QY 1066 TATGTCAGCTATGTTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1125
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 14493 GTGTGTGTTTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 14552

QY 1126 CGAATTCACCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1185
```

```
DB 14553 TGTGTGTTTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 14612
QY 1186 GTAAGTGTGTGACTGTAAGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1225
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 14613 GTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 14652
```

RESULT 26  
AL606962 223854 bp DNA Linear ROD 20-JUN-2002  
LOCUS  
DEFINITION Mouse DNA sequence from clone RP23-29H22 on chromosome 4, complete sequence.  
ACCESSION AL606962  
VERSION AL606962.22 GI:21537443  
KEYWORDS HTG.  
SOURCE house mouse.  
ORGANISM Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
1 (bases 1 to 223854)  
REFERENCE Babage,A.  
TITLE Direct Submission  
JOURNAL Submitted (20-JUN-2002) Wellcome Trust Sanger Institute, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk  
COMMENT On Jun 21, 2002 this sequence version replaced gi:21530967.  
----- Genome Center  
Center: Wellcome Trust Sanger Institute  
Center code: SC  
Web site: <http://www.sanger.ac.uk>  
Contact: [humquery@sanger.ac.uk](mailto:humquery@sanger.ac.uk)

During sequence assembly data is compared from overlapping clones. Where differences are found these are annotated as variations together with a note of the overlapping clone name. Note that the variation annotation may not be found in the sequence submission corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above.

This sequence was finished as follows unless otherwise noted: all regions were either double-stranded or sequenced with an alternate chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one plasmid subclone or more than one M3 subclone; and the assembly was confirmed by restriction digest. The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: Em., EMBL; Sw., SWISSPROT; Tr., TREMBL; Wp., WORMPEP; information on the WORMPEP database can be found at [http://www.sanger.ac.uk/Projects/C\\_elegans/wormpep](http://www.sanger.ac.uk/Projects/C_elegans/wormpep) RP23-29H22 is from the RPCI-23 Mouse PAC Library

constructed by the group of Pieter de Jong.  
For further details see <http://www.chori.org/bacpac/home.htm>  
VECTOR: pBAC3.6.

FEATURES  
source 1..223854  
/organism="Mus musculus"  
/db\_xref="taxon:10090"  
/chromosome="4"  
/clone="RP23-29H22"  
/clone\_1lb="RPCI-23"  
BASE COUNT 57553 a 50381 c 52017 g 63903 t  
ORIGIN

Query Match 3.8%; Score 56; DB 10; Length 223854;  
Best Local Similarity 47.9%; Pred. No. 8.3e-05;  
Matches 204; Conservative 0; Mismatches 210; Indels 12; Gaps 1;

```
QY 664 TAAGTACATGACGACGATATTCGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 723
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 97212 TAAGTCAAAATTTGTAATCTTCTCTGCGGTGTGTGTGTGTGTGTGTGTGTGTGT 97271
```

Oy 724 CTACCTGATCCGGTGTGTGTATGTGTGACAGCCCTTGATATAAACAATGTTGGATGA 783  
 Db 97272 GTGTGTGATGTATGTATGTGTGATGCGTGTGATGTGTGTGTGTGTATGTATGTCTGT 97331  
 Oy 768 CTTTGTGACAGTTCAGTAGACAGATGATTCATTCGCGTCAATGTGTGTGTGTGTGTGAG 843  
 Db 97332 ATGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTG 97391  
 Oy 844 GTGACGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTG 903  
 Db 97332 TGTATGATGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 97451  
 Oy 904 TGATTGAGACCACTTGAGAGAGAGAACTCATGAATAATCTATGCAACGATGGAAGTG 963  
 Db 97452 TGAGTGTATGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 97502  
 Oy 964 TGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1023  
 Db 97503 --TATGTGTATGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 97559  
 Oy 1024 TGGAGAGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1083  
 Db 97560 TGATGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 97619  
 Oy 1084 GTGGGT 1089  
 Db 97620 GTGTGT 97625

RESULT 27  
 AC111443/C  
 LOCUS  
 DEFINITION Rattus norvegicus clone CH230-192H9, \*\*\* SEQUENCING IN PROGRESS  
 AC111443 171870 bp DNA linear HTG 13-JUL-2002  
 Rattus norvegicus  
 \*\*\* 68 unordered pieces.  
 AC111443  
 VERSION AC111443.2 GI:21735991  
 KEYWORDS HTG; HTGS. PHASE1.  
 SOURCE Norway rat.  
 ORGANISM Rattus norvegicus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;  
 Rattus.  
 1 (bases 1 to 171870)  
 Murny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C.,  
 Albrooks,S.L., Amaralunge,H.C., Are,J.R., Ayala,M., Banks,T.,  
 Barbata,J., Benton,J., Blinze,K., Blankenburg,K., Bonin,D.,  
 Bouck,J., Bowle,S., Brileva,M., Brown,E., Brown,M., Bryant,N.P.,  
 Bunay,C., Burch,P., Burkett,C., Burrell,K.D., Byrd,N.C.,  
 Caron,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D.,  
 Chen,G., Chen,R., Chen,Z., Chowdhry,I., Christopoulos,C.,  
 Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R.,  
 Davis,M.L., Davis,C., Davy-Carroll,L., Dederich,D.A.,  
 Delaney,K.R., Delgado,O., Dem,A.L., Ding,Y., Dinh,H.H.,  
 Douhaite,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.J.,  
 Earhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Escotto,M.,  
 Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Frantz,P.,  
 Gabriel,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R.,  
 Gorrell,D.H., Guevara,W., Gunaratne,P., Hale,S., Hamilton,K.,  
 Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A., Hernandez,J.,  
 Hernandez,O., Hodgson,A., Hogues,M., Hollway,C., Hollins,B.,  
 Homel,F., Howard,S., Huber,J., Hulik,S., Hume,J., Jackson,L.E.,  
 Jacobson,B., Jia,Y., Johnson,R., Jollivet,S., Joudah,S.,  
 Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J., Kovar,C.,  
 Kravovic,J., Kureshi,A., Landy,N., Leal,B., Lewis,L.C., Lewis,L.,  
 Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W., Louised,H.,  
 Lozano,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R., Ma,J.,  
 Mareshwari,M., Mapua,P., Martin,R., Martindale,A., Martinez,E.,  
 Maesey,E., Mawhinney,E., McLeod,M.P., Meador,M., Mei,G., Metzger,M.,  
 Moser,G., Miner,Z., Mitchell,T., Mohabbat,K., Morgan,M., Morris,S.,  
 Moser,M., Neal,D., Newton,J., Newton,N., Nguyen,A., Nguyen,N.,  
 Nguyen,N., Nickerson,E., Nwokenko,S., Ogun,M., Okunoluwa,G.,  
 Oregunye,N., Oviedo,R., Pace,A., Payton,B., Peery,J., Perez,L.,  
 Peters,L., Pickens,R., Primus,E., Pu,L.L., Qules,M., Ren,Y.,

TITLE  
 JOURNAL  
 REFERENCE  
 AUTHORS  
 TITLE  
 JOURNAL

REFERENCE  
 AUTHORS  
 TITLE  
 JOURNAL

COMMENT

Rives,M., Rojars,A., Rojokkan,I., Rolfe,M., Ruiz,S., Savary,G.,  
 Scherer,S., Scott,G., Shen,H., Shoochari,N., Sisson,I.,  
 Sodergren,E., Sotak,T., Sparks,A., Stanley,H., Stone,H.,  
 Sutton,A., Syarik,A., Taber,P., Tameria,A., Tamerias,K., Tang,H.,  
 Tansey,J., Taylor,C., Taylor,T., Telford,B., Thomas,N., Thomas,S.,  
 Umanai,K., Vasquez,L., Vera,V., Villalon,D., Vinson,R., Wang,Q.,  
 Wang,S., Ward-Moore,S., Warren,R., Washington,C., Watlington,S.,  
 Williams,G., Williamson,A., Wiczek,R., Wooden,S., Worley,K.,  
 Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D.,  
 Weinstein,G., and Gibbs,R.  
 Direct Submission  
 Unpublished  
 2 (bases 1 to 171870)  
 Morley,K.C.  
 Direct Submission  
 Submitted (19-FEB-2002) Human Genome Sequencing Center, Department  
 of Molecular and Human Genetics, Baylor College of Medicine, One  
 Baylor Plaza, Houston, TX 77030, USA  
 3 (bases 1 to 171870)  
 Morley,K.C.  
 Direct Submission  
 Submitted (13-JUL-2002) Human Genome Sequencing Center, Department  
 of Molecular and Human Genetics, Baylor College of Medicine, One  
 Baylor Plaza, Houston, TX 77030, USA  
 On Jul 12, 2002 this sequence version replaced gi:18701208.  
 -----  
 Center: Baylor College of Medicine  
 Genome Center  
 -----  
 Center code: BCM  
 Web site: <http://www.hgsc.bcm.tmc.edu/>  
 Contact: [hgsc-help@bcm.tmc.edu](mailto:hgsc-help@bcm.tmc.edu)  
 -----  
 Project Information  
 Center project name: GMIO  
 Center clone name: CH230-192H9  
 -----  
 Summary Statistics  
 Sequencing vector: Plasmid;  
 Chemistry: Dye-terminator Big Dye 100% of reads  
 Assembly program: Phrap; version 0.990329  
 Consensus quality: 109309 bases at least Q40  
 Consensus quality: 113392 bases at least Q30  
 Consensus quality: 116866 bases at least Q20  
 -----  
 \* NOTE: Estimated insert size may differ from sequence length.  
 \* (see [http://www.hgsc.bcm.tmc.edu/docs/genbank\\_draft\\_data.html](http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html)).  
 \* NOTE: This is a 'working draft' sequence. It currently  
 \* consists of 68 contigs. The true order of the pieces  
 \* is not known and their order in this sequence record is  
 \* arbitrary. Gaps between the contigs are represented as  
 \* runs of N, but the exact sizes of the gaps are unknown.  
 \* This record will be updated with the finished sequence  
 \* as soon as it is available and the accession number will  
 \* be preserved.  
 1 1047: contig of 1047 bp in length  
 \* 1048 1147: gap of unknown length  
 \* 1148 2592: contig of 1445 bp in length  
 \* 2593 2692: gap of unknown length  
 \* 2693 3969: contig of 1277 bp in length  
 \* 3970 4063: gap of unknown length  
 \* 4063 5194: contig of 1125 bp in length  
 \* 5195 5294: gap of unknown length  
 \* 5294 6671: contig of 1377 bp in length  
 \* 6671 6771: gap of unknown length  
 \* 6771 7929: contig of 1158 bp in length  
 \* 7929 7930: gap of unknown length  
 \* 7930 8030: gap of unknown length  
 \* 8030 9231: contig of 1202 bp in length  
 \* 9231 9332: gap of unknown length  
 \* 9332 10939: contig of 1608 bp in length  
 \* 10940 11039: gap of unknown length  
 \* 11040 12627: contig of 1587 bp in length  
 \* 12627 12725: gap of unknown length  
 \* 12725 14176: contig of 1450 bp in length  
 \* 14176 14276: gap of unknown length  
 \* 14276 15463: contig of 1187 bp in length  
 \* 15463 15563: gap of unknown length

```

* 15564 17002: contig of 1439 bp in length
* 17003 17102: gap of unknown length
* 17003 18113: contig of 1211 bp in length
* 18314 18413: gap of unknown length
* 18414 19987: contig of 1574 bp in length
* 19988 20087: gap of unknown length
* 20088 21888: contig of 1801 bp in length
* 21889 21988: gap of unknown length
* 21989 23279: contig of 1291 bp in length
* 23280 23380: gap of unknown length
* 23380 24592: contig of 1213 bp in length
* 24593 24692: gap of unknown length
* 24693 26165: contig of 1473 bp in length
* 26166 26265: gap of unknown length
* 26266 27781: contig of 1516 bp in length
* 27782 27881: gap of unknown length
* 27882 29053: contig of 1172 bp in length
* 29054 29153: gap of unknown length
* 29154 30241: contig of 1088 bp in length
* 30242 30341: gap of unknown length
* 30342 31854: contig of 1513 bp in length
* 31855 31954: gap of unknown length
* 31955 33832: contig of 1878 bp in length
* 33833 33932: gap of unknown length
* 33933 35770: contig of 1838 bp in length
* 35771 35870: gap of unknown length
* 35871 37519: contig of 1648 bp in length
* 37519 37618: gap of unknown length
* 37619 39608: contig of 1990 bp in length
* 39609 39708: gap of unknown length
* 39709 41149: contig of 1441 bp in length
* 41150 41249: gap of unknown length
* 41250 43176: contig of 1927 bp in length
* 43177 43276: gap of unknown length
* 43277 44621: contig of 1545 bp in length
* 44622 44921: gap of unknown length
* 44922 47202: contig of 2281 bp in length
* 47203 47302: gap of unknown length
* 47303 48414: contig of 1112 bp in length
* 48415 48514: gap of unknown length
* 48515 50405: contig of 1891 bp in length
* 50406 50505: gap of unknown length
* 50506 52132: contig of 1627 bp in length
* 52133 52232: gap of unknown length
* 52233 53510: contig of 1278 bp in length
* 53511 53756: gap of unknown length
* 53757 55756: contig of 2146 bp in length
* 55757 55856: gap of unknown length
* 55857 57761: contig of 1905 bp in length
* 57762 57861: gap of unknown length
* 57862 59772: contig of 1911 bp in length
* 59773 59872: gap of unknown length
* 59873 62180: contig of 2308 bp in length
* 62181 62280: gap of unknown length
* 62281 64314: contig of 2034 bp in length
* 64315 64414: gap of unknown length
* 64415 66757: contig of 2343 bp in length
* 66758 66857: gap of unknown length
* 66858 69014: contig of 2157 bp in length
* 69015 69114: gap of unknown length
* 69115 71354: contig of 2240 bp in length
* 71355 71454: gap of unknown length
* 71455 73566: contig of 2112 bp in length
* 73567 73666: gap of unknown length
* 73667 75603: contig of 1837 bp in length
* 75604 77337: gap of unknown length
* 77338 77437: gap of unknown length
* 77438 81100: contig of 3663 bp in length
* 81101 81200: gap of unknown length
* 81201 84069: contig of 2869 bp in length
* 84070 84169: gap of unknown length
* 84170 86014: contig of 1845 bp in length

```

Query Match 3.8%; Score 55.8; DB 2; Length 171870;  
 Best Local Similarity 47.1%; Pred. No. 9.3e-05;  
 Matches 171; Conservative 0; Mismatches 192; Indels 0; Gaps 0;

```

* 86015 86114: gap of unknown length
* 86115 88884: contig of 2770 bp in length
* 88885 88984: gap of unknown length
* 88985 92923: contig of 3939 bp in length
* 92924 93023: gap of unknown length
* 93024 95634: contig of 2611 bp in length
* 95635 95734: gap of unknown length
* 95735 98938: contig of 3104 bp in length
* 98939 103775: contig of 4837 bp in length
* 103776 103875: gap of unknown length
* 103876 108073: contig of 4198 bp in length

QY 673 TGTCACTAGTATTCTGTGTCTGTGTATCTGTTGTGTCTGTCTGTCTACTTGTAT
Db 121365 TGTGTGTAGTGTGTGTATGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTG 121306

QY 733 CCGGTGTGTATTTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTG 792
Db 121305 TGAAGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTG 121246

QY 793 AGTTCAGTAGCAGAGTGTATTCATCTCGTCAATGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTG 852
Db 121245 TGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTG 121186

QY 853 TGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTG 912
Db 121185 AGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTG 121126

QY 913 CCACCTTGAGAGAGAGAAGACTCATGAATATCTATGACAAACATGGAAGTGTGTGTGTGTGTG 972
Db 121125 TGAATGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTG 121066

QY 973 CATGAAGTAGGGGGTCAATGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTG 1032
Db 121065 TGAATGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTG 121006

QY 1033 ACT 1035
Db 121005 AGT 121003

```

RESULT 28  
 AL671970/c 211969 bp DNA linear ROD 26-JUN-2002  
 LOCUS Mouse DNA sequence from clone RP23-152A20 on chromosome 4, complete  
 DEFINITION sequence.  
 ACCESSION AL671970  
 VERSION AL671970.7 GI:21531211  
 KEYWORDS HTG.  
 SOURCE house mouse.  
 ORGANISM Mus musculus.  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 1 (bases 1 to 211969)  
 REFERENCE Blakey, S.  
 Direct Submission  
 Submitted (26-JUN-2002) Wellcome Trust Sanger Institute, Hinxton,  
 Cambridgeshire, CB10 1SA, UK. E-mail enquiries:  
 humquerry@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk  
 On Jun 21, 2002 this sequence version replaced gi:21212233.  
 COMMENT ----- Genome Center  
 Center: Wellcome Trust Sanger Institute  
 Center code: SC  
 Web site: http://www.sanger.ac.uk  
 Contact: humquerry@sanger.ac.uk  
 -----

During sequence assembly data is compared from overlapping clones.  
 Where differences are found these are annotated as variations











* be preerred	1	1117:	contig of 1117 bp in length
*	1118	1217:	gap of unknown length
*	1218	2748:	contig of 1531 bp in length
*	2749	2848:	gap of unknown length
*	2849	4131:	contig of 1283 bp in length
*	4132	4231:	gap of unknown length
*	4232	5848:	contig of 1617 bp in length
*	5849	5948:	gap of unknown length
*	5949	7612:	contig of 1664 bp in length
*	7613	7712:	gap of unknown length
*	7713	9102:	contig of 1390 bp in length
*	9103	9202:	gap of unknown length
*	9203	10347:	contig of 1145 bp in length
*	10348	10447:	gap of unknown length
*	10449	11490:	contig of 1043 bp in length
*	11491	11590:	gap of unknown length
*	11591	12762:	contig of 1172 bp in length
*	12763	12862:	gap of unknown length
*	12863	14281:	contig of 1419 bp in length
*	14282	14381:	gap of unknown length
*	14382	15962:	contig of 1581 bp in length
*	15963	16062:	gap of unknown length
*	16063	17128:	contig of 1066 bp in length
*	17129	17228:	gap of unknown length
*	17229	18437:	contig of 1209 bp in length
*	18438	18537:	gap of unknown length
*	18538	19708:	contig of 1171 bp in length
*	19709	19808:	gap of unknown length
*	19809	21023:	contig of 1215 bp in length
*	21024	21123:	gap of unknown length
*	21124	22783:	contig of 1660 bp in length
*	22784	22883:	gap of unknown length
*	22884	24459:	contig of 1576 bp in length
*	24460	24559:	gap of unknown length
*	24560	26519:	contig of 1960 bp in length
*	26520	26619:	gap of unknown length
*	26620	27797:	contig of 1178 bp in length
*	27798	27897:	gap of unknown length
*	27898	29101:	contig of 1204 bp in length
*	29102	29201:	gap of unknown length
*	29202	31713:	contig of 2512 bp in length
*	31714	31813:	gap of unknown length
*	31814	33291:	contig of 1478 bp in length
*	33292	33391:	gap of unknown length
*	33392	34642:	contig of 1251 bp in length
*	34643	34742:	gap of unknown length
*	34743	36339:	contig of 1597 bp in length
*	36340	36439:	gap of unknown length
*	36440	38764:	contig of 2325 bp in length
*	38765	38864:	gap of unknown length
*	38865	40105:	contig of 1241 bp in length
*	40106	40205:	gap of unknown length
*	40206	41681:	contig of 1476 bp in length
*	41682	41781:	gap of unknown length
*	41782	43467:	contig of 1686 bp in length
*	43468	43567:	gap of unknown length
*	43568	45737:	contig of 2170 bp in length
*	45738	45837:	gap of unknown length
*	45838	47359:	contig of 1522 bp in length
*	47360	47459:	gap of unknown length
*	47460	49072:	contig of 1613 bp in length
*	49073	49172:	gap of unknown length
*	49173	50987:	contig of 1815 bp in length
*	50988	51087:	gap of unknown length
*	51088	52880:	contig of 1793 bp in length
*	52881	55940:	gap of unknown length
*	55941	55944:	contig of 2064 bp in length
*	55945	55144:	gap of unknown length
*	55145	57633:	contig of 2489 bp in length
*	57634	57733:	gap of unknown length
*	57734	59322:	contig of 1589 bp in length
*	59323	59322:	gap of unknown length

*	59423	61121: contig of 1699 bp in length
*	61122	61121: gap of unknown length
*	61122	63639: contig of 2418 bp in length
*	63340	63739: gap of unknown length
*	63740	66110: contig of 2371 bp in length
*	66111	66210: gap of unknown length
*	66211	68352: contig of 2142 bp in length
*	68353	68452: gap of unknown length
*	68453	71146: contig of 2634 bp in length
*	68453	71147: gap of unknown length
*	71147	74073: contig of 2827 bp in length
*	74074	74173: gap of unknown length
*	74174	76170: contig of 1997 bp in length
*	76171	76270: gap of unknown length
*	76271	78912: contig of 2642 bp in length
*	78913	79012: gap of unknown length
*	79013	81419: contig of 2407 bp in length
*	81519	81519: gap of unknown length
*	81520	83366: contig of 1847 bp in length
*	83367	83466: gap of unknown length
*	83467	86240: contig of 2774 bp in length
*	86241	86340: gap of unknown length
*	86341	89381: contig of 3431 bp in length
*	89382	89391: gap of unknown length
*	89392	92650: contig of 2719 bp in length
*	92651	92750: gap of unknown length
*	92751	95640: contig of 2880 bp in length
*	95641	95740: gap of unknown length
*	95741	99221: contig of 3481 bp in length
*	99222	99231: gap of unknown length
*	99322	10259: contig of 3278 bp in length
*	102600	102639: gap of unknown length
*	102700	105946: contig of 3247 bp in length
*	105947	106046: gap of unknown length
*	106047	106978: contig of 3739 bp in length

Query Match	3.8%	Score 55.4;	DB 2;	Length 185413;
Best Local Similarity	44.6%	Pred. No. 0.00012;		
Matches 255; Conservative	0;	Mismatches 312;	Indels 5;	Gaps 2;

[illegible]

Db 7504 NTGGGGGGTTTGTGTTTGGGTTTGTGTTTGTGTTGTTGGGGTGTGGG 7563  
 Oy 1195 TGTACTGACTGTTGTGTGTTGTTGTTGTTG 1226  
 Db 7564 TTTTGTGGTGTGTTTGTGTTGTTGTTGTTG 7595

## RESULT 32

AC114869

LOCUS Rattus norvegicus clone CH230-290G24, \*\*\* SEQUENCING IN PROGRESS

DEFINITION \*\*\* 67 unordered pieces.

ACCESSION AC114869

VERSION AC114869.2 GI:21745661

KEYWORDS HTG; HTGS; PHASE1.

SOURCE Norway rat.

ORGANISM Rattus norvegicus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;

## REFERENCE

AUTHORS

1 (bases 1 to 188030)  
 Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C.,  
 Alsbrooks,S.L., Amaralunge,H.C., Are,J.R., Ayale,M., Banks,T.,  
 Barbata,J., Benton,J., Bimaga,K., Blankenburg,K., Bonnin,D.,  
 Bouck,J., Bowie,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P.,  
 Buhay,C., Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C.,  
 Carron,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D.,  
 Chen,G., Chen,R., Chen,Z., Chowdhry,I., Christopoulos,C.,  
 Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R.,  
 Delaney,K.R., Delgado,O., Denn,A.L., Ding,Y., Dinh,H.H.,  
 Douthwaite,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.J.,  
 Earnhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Escotto,M.,  
 Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Frantz,P.,  
 Gabisi,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R.,  
 Gorell,J.H., Guevara,W., Gunaratne,P., Hale,S., Hamilton,K.,  
 Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A., Hernandez,J.,  
 Hernandez,O., Hodgson,A., Hognes,M., Holloway,C., Hollins,B.,  
 Homi,F., Howard,S., Huber,J., Hulyk,S., Hume,J., Jackson,L.E.,  
 Jacobson,B., Jia,Y., Johnson,R., Jolivet,S., Joudan,S.,  
 Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J., Kovar,C.,  
 Kravovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C., Lewis,L.,  
 Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W., Louieged,H.,  
 Lozdo,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R., Ma,J.,  
 Maheshwari,M., Mapua,P., Martin,R., Martindale,A., Martinez,R.,  
 Maseley,E., Mawhney,E., McLeod,M.P., Meador,M., Mei,G., Metzker,M.,  
 Miner,G., Miner,Z., Mitchell,T., Mohabbat,K., Morgan,M., Morris,S.,  
 Moser,M., Neal,D., Newton,J., Newton,N., Nguyen,A., Nguyen,N.,  
 Nguyen,N., Nickerson,E., Nwokenkwo,S., Oguh,M., Okunolu,G.,  
 Oragunye,N., Oviedo,R., Pace,A., Payton,B., Peery,J., Perez,L.,  
 Peters,L., Pichens,R., Primus,E., Pu,L., Quiles,M., Ren,Y.,  
 Rivers,M., Rojas,A., RojudoKan,I., Rolfe,M., Ruiz,S., Savery,G.,  
 Scherer,S., Scott,G., Shen,H., Shoshari,N., Sisson,I.,  
 Sudergen,E., Sonaike,T., Sparks,A., Stanley,H., Stone,H.,  
 Sutton,A., Svatek,A., Tabor,P., Tamerisa,A., Tamerisa,K., Tang,H.,  
 Taney,J., Taylor,C., Taylor,T., Telford,B., Thomas,N., Thomas,S.,  
 Umami,K., Vaequez,L., Vera,V., Villalon,D., Vinson,R., Wang,Q.,  
 Wang,S., Ward-Moore,S., Warren,R., Washington,C., Watlington,S.,  
 Williams,G., Williamson,A., Wleczky,R., Wooden,S., Worley,K.,  
 Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D.,  
 Weinstein,G. and Gibbs,R.  
 Direct Submission  
 Unpublished  
 2 (bases 1 to 188030)  
 Worley,K.C.  
 Direct Submission  
 Submitted (12-04-2002) Human Genome Sequencing Center, Department  
 of Molecular and Human Genetics, Baylor College of Medicine, One  
 Baylor Plaza, Houston, TX 77030, USA  
 3 (bases 1 to 188030)  
 Worley,K.C.  
 Direct Submission

## JOURNAL

## COMMENT

Submitted (17-JUL-2002) Human Genome Sequencing Center, Department  
 of Molecular and Human Genetics, Baylor College of Medicine, One  
 Baylor Plaza, Houston, TX 77030, USA  
 On Jul 14, 2002 this sequence version replaced gi:19352257.

----- Genome Center  
 Center: Baylor College of Medicine

Center code: BCM  
 Web site: <http://www.hgsc.bcm.tmc.edu/>  
 Contact: [hgsc-help@bcm.tmc.edu](mailto:hgsc-help@bcm.tmc.edu)

----- Project Information  
 Center project name: GSYH  
 Center clone name: CH230-290G24

----- Summary Statistics  
 Sequencing vector: Plasmid;  
 Chemistry: Dye-terminator Big Dye; 100% of reads  
 Assembly program: Phrap; version 0.990329

Consensus quality: 138957 bases at least Q40  
 Consensus quality: 145134 bases at least Q30  
 Consensus quality: 150777 bases at least Q20

-----  
 \* NOTE: Estimated insert size may differ from sequence length  
 \* (see [http://www.hgsc.bcm.tmc.edu/docs/genbank\\_draft\\_data.html](http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html)).  
 \* NOTE: This is a 'working draft' sequence. It currently  
 \* consists of 67 contigs. The true order of the pieces is  
 \* is not known and their order in this sequence record is  
 \* arbitrary. Gaps between the contigs are represented as  
 \* runs of N, but the exact sizes of the gaps are unknown.  
 \* This record will be updated with the finished sequence  
 \* as soon as it is available and the accession number will  
 \* be preserved.

1 1544: contig of 1544 bp in length  
 1545 1644: gap of unknown length  
 1645 2662: contig of 1018 bp in length  
 2663 2762: gap of unknown length  
 2763 3797: contig of 1035 bp in length  
 3798 3897: gap of unknown length  
 3898 5351: contig of 1454 bp in length  
 5352 5452: gap of unknown length  
 5453 6657: contig of 1206 bp in length  
 6658 6757: gap of unknown length  
 6758 7838: contig of 1081 bp in length  
 7839 7938: gap of unknown length  
 7939 9039: contig of 1101 bp in length  
 9040 9139: gap of unknown length  
 9140 10371: contig of 1232 bp in length  
 10372 10471: gap of unknown length  
 10472 12261: contig of 1790 bp in length  
 12262 12361: gap of unknown length  
 12362 13723: contig of 1362 bp in length  
 13724 13823: gap of unknown length  
 13824 14910: contig of 1087 bp in length  
 14911 15010: gap of unknown length  
 15011 16072: contig of 1062 bp in length  
 16073 16172: gap of unknown length  
 16173 17803: contig of 1631 bp in length  
 17804 17903: gap of unknown length  
 17904 20002: contig of 2099 bp in length  
 20003 20102: gap of unknown length  
 20103 21780: contig of 1678 bp in length  
 21781 21880: gap of unknown length  
 21881 23061: contig of 1181 bp in length  
 23062 23161: gap of unknown length  
 23162 25413: contig of 2252 bp in length  
 25414 25513: gap of unknown length  
 25514 27401: contig of 1868 bp in length  
 27402 27501: gap of unknown length  
 27502 28839: contig of 1338 bp in length  
 28840 28939: gap of unknown length  
 28940 31202: contig of 2263 bp in length  
 31203 31302: gap of unknown length  
 31303 32511: contig of 1209 bp in length  
 32512 32611: gap of unknown length  
 32612 34064: contig of 1453 bp in length

*	34065	34164:	gap of unknown length
*	34165	36113:	contig of 1949 bp in length
*	36114	36213:	gap of unknown length
*	36214	38441:	contig of 2228 bp in length
*	38442	38541:	gap of unknown length
*	38542	40593:	contig of 2052 bp in length
*	40594	40693:	gap of unknown length
*	40694	43224:	contig of 2531 bp in length
*	43325	43324:	gap of unknown length
*	43325	45302:	contig of 1978 bp in length
*	45303	45402:	gap of unknown length
*	45403	46693:	contig of 1291 bp in length
*	46694	46793:	gap of unknown length
*	46794	48075:	contig of 1282 bp in length
*	48076	48175:	gap of unknown length
*	48176	50357:	contig of 2182 bp in length
*	50358	50457:	gap of unknown length
*	50458	52659:	contig of 2202 bp in length
*	52660	52759:	gap of unknown length
*	52760	55935:	contig of 3176 bp in length
*	55936	56035:	gap of unknown length
*	56036	59013:	contig of 2978 bp in length
*	59014	59113:	gap of unknown length
*	59114	60613:	contig of 1500 bp in length
*	60614	60713:	gap of unknown length
*	60714	62959:	contig of 2246 bp in length
*	62960	63059:	gap of unknown length
*	63060	65488:	contig of 2429 bp in length
*	65489	65589:	gap of unknown length
*	65589	67416:	contig of 1828 bp in length
*	67417	67516:	gap of unknown length
*	67517	71482:	contig of 3966 bp in length
*	71483	71582:	gap of unknown length
*	71583	73885:	contig of 2303 bp in length
*	73886	73985:	gap of unknown length
*	73986	77240:	contig of 3255 bp in length
*	77241	77340:	gap of unknown length
*	77341	79542:	contig of 2202 bp in length
*	79543	79642:	gap of unknown length
*	79643	82075:	contig of 2433 bp in length
*	82076	82175:	gap of unknown length
*	82176	85936:	contig of 3761 bp in length
*	85937	86036:	gap of unknown length
*	86037	89476:	contig of 3440 bp in length
*	89477	89576:	gap of unknown length
*	89577	92798:	contig of 3222 bp in length
*	92799	92898:	gap of unknown length
*	92899	95595:	contig of 2697 bp in length
*	95596	95695:	gap of unknown length
*	95696	97631:	contig of 1936 bp in length
*	97632	97731:	gap of unknown length
*	97732	100455:	contig of 2724 bp in length
*	100456	100555:	gap of unknown length
*	100556	104034:	contig of 3479 bp in length
*	104035	104134:	gap of unknown length
*	104135	107314:	contig of 3180 bp in length
*	107315	107414:	gap of unknown length
*	107415	111566:	contig of 4152 bp in length
*	111567	111666:	gap of unknown length
*	111667	114264:	contig of 2598 bp in length
*	114265	114364:	gap of unknown length
*	114365	117509:	contig of 3145 bp in length
*	117510	117609:	gap of unknown length
*	117610	120732:	contig of 3123 bp in length

[illegible]

**RESULT 33**

**Locus AC110512/c**

**DEFINITION** Mus musculus clone RP24-134E9, WORKING DRAFT SEQUENCE, 13 ordered pieces.

**ACCESSION** AC110512

**VERSION** AC110512.4 GI:22381906

**KEYWORDS** HTG, HTGS\_PHASE2, HTGS\_DRAFT, HTGS\_FULFUP.

**SOURCE** house mouse.

**ORGANISM** Mus musculus  
EuKaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Scurionathii; Muridae; Murinae; Mus.

**REFERENCE** 1 (bases 1 to 167892)  
**AUTHORS** Birren,B., Linton,L., Nusbaum,C. and Lander,E.  
**JOURNAL** MUs musculur, Clone RP24-134E9  
**TITLE** Unpublished  
**REFERENCE** 2 (bases 1 to 167892)

Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N., Anderson,S., Barua,N., Bastien,V., Boguslavsky,L., Bouckgaler,B., Brown,A., Camarata,J., Campioiano,A., Chang,J., Charzo,R., Chepel,Y., Colangelo,M., Collins,S., Collimore,A., Cook,A., Cooke,P., Dekrellano,K., Dewar,K., Diaz,J.S., Dodge,S., Faro,S., Ferreira,P., Fitzhugh,W., Gage,D., Galgan,J., Gardyna,S., Glnde,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N., Hagos,B., Hooton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C., Kanat,A., Karatas,A., Kells,C., LaRoque,K., Lamazars,R., Landers,T., Lehoczy,J., Levine,R., Liu,G., Maclean.C., MacDonald,P., Major,J., Marguis,N., Matthews,C., McCarthy,M., Mcowan,P., McKernan,K., Meldrim.J., Meheus,L., Mihova,T., Menga,V., Murphy,T., Naylor,J., Nguyen,C., Nicol.R., Norbu,C., Norman,C.H., O'Connor.T., O'Donnell.P., O'Neil.D., Oliver,J., Peterson.K., Phunhang,P., Pierre.N., Pollara.V., Raymond.C., Retta.R., Rieback.M., Riley,R., Rise,C., Rogov,P., Roman.J., Rosetti,M., Roy.A., Santos,R., Schauer,S., Schnpack,R., Seaman,S., Severly,P., Spencer.B., Stange-Thomann,N., Stojanovic,N., Strauss,N., Subramanian.A., Talmes,J., Teefaye,S., Theodore,J., Topham,K., Travers,M., Travis,N., Trigilio,J., Vassiliy,H., Viel,R., Vo.A., Wilson,B., Wu,X., Wyman,D., Ye.W.J., Young.G., Zainoun,J., Zemcek,L., Zimmer,A. and Zody,M.

**TITLE** Direct Submission



```

|||||
Db 106912 TGAGGTCAAACTGACGTCGATGTCATGACAGTTGACATGTCGTGATATC 106853
Oy 1097 AGCGTGATATGTCGCGATGATATTTTCGAATTCAGTTGGTTGTTGTTTC 1156
Db 106852 GGTACGGGGAACATGAGAGTCTCTCTCTCTTTCACCTTTGTTGTTTGTG 106793
Oy 1157 GTTGTGTTTGTCTTCAT 1178
Db 106792 TTTGTTTTTGTCTTCACCT 106771

RESULT 34
AC109062 176192 bp DNA linear HTG 13-JUL-2002
AC109062 Rattus norvegicus clone CH230-96P15, *** SEQUENCING IN PROGRESS
LOCUS *** 61 unordered pieces.
DEFINITION
AC109062
AC109062 AC109062.3 GI:21737853
VERSION HTG: HTGS PHASE1.
KEYWORDS
SOURCE Norway rat.
ORGANISM Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
1 (bases 1 to 176192)
Murny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C.,
Albrooke,S.L., Amaralunge,H.C., Are,J.R., Ayale,M., Banks,T.,
Babaria,J., Benton,J., Blincoe,K., Blankenburg,K., Bonini,D.,
Bouck,J., Bowle,S., Bileva,M., Brown,E., Brown,M., Bryant,N.P.,
Bunay,K., Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C.,
Carroll,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D.,
Chen,G., Chen,R., Chen,Z., Chodhury,I., Christopoulos,C.,
Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R.,
Dayala,M.L., Davis,C., Davy-Carroll,L., Dederich,D.A.,
DeJanel,K.R., Delgado,O., Dem,A.L., Ding,Y., Dinh,H.H.,
Duchateau,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.J.,
Eamhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Escotto,M.,
Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Frantz,P.,
Gabriel,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R.,
Gorell,U.H., Guevara,W., Gunaratne,P., Hale,S., Hamilton,K.,
Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A., Hernandez,J.,
Hernandez,O., Hodgson,A., Hognes,M., Hollway,C., Hollins,B.,
Homsli,F., Howard,S., Huber,J., Hulik,S., Hume,J., Jackson,L.E.,
Jacobson,B., Jia,Y., Johnson,R., Jolivet,S., Joudh,S.,
Karlsom,E., Kelly,S., Khan,U., King,L., Korah,U., Kovar,C.,
Kratovic,J., Kureshi,A., Landy,N., Leal,B., Lewis,L.C., Lewis,L.,
L.J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W., Louised,H.,
Lozano,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R., Ma,J.,
Maheshwari,M., Mapua,E., McLeod,M.P., Meador,M., Mei,G., Metzger,M.,
Miner,G., Miner,Z., Mitchell,T., Mohabbat,K., Morgan,M., Morris,S.,
Moer,N., Neal,D., Newton,J., Newton,N., Nguyen,A., Nguyen,N.,
Nguyen,N., Nickerson,E., Nwokweto,S., Ogun,M., Okunolu,G.,
Oregunye,N., Oviedo,R., Pace,A., Payton,B., Peary,J., Perez,L.,
Peterson,L., Pickens,R., Primus,E., Pu,L.L., Quiles,M., Ren,Y.,
Rives,M., Rojas,A., Rojokan,I., Rolfe,M., Ruiz,S., Savary,G.,
Scherrer,S., Scott,G., Shen,H., Shoostari,N., Sisoon,I.,
Soederren,E., Sonalke,T., Sparks,A., Stanley,H., Stone,H.,
Sutton,A., Svatek,A., Taber,P., Tamerisa,A., Tamerisa,K., Tang,H.,
Tansey,J., Taylor,C., Taylor,T., Telford,B., Thomas,N., Thomas,S.,
Umanu,K., Vasquez,L., Vera,V., Villalon,D., Vinson,R., Wang,O.,
Wang,S., Ward-Moore,S., Warren,R., Washington,C., Watlington,S.,
Williams,G., Williamson,A., Wiczek,R., Wooden,S., Worley,K.,
Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zortilla,S., Nelson,D.,
Weinstock,G., and Gibbs,R.
Unpublished
Direct Submission
2 (bases 1 to 176192)
Worley,K.C.
TITLE
JOURNAL
REFERENCE
AUTHORS
JOURNAL
Submitted (03-FEB-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One

```

```

REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT
Baylor Plaza, Houston, TX 77030, USA
3 (bases 1 to 176192)
Worley,K.C.
Direct Submission
Submitted (13-JUL-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
On Jul 12, 2002 this sequence version replaced gi:18846790.
----- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: http://www.hgsc.bcm.tmc.edu/
Contact: hgsc-help@bcm.tmc.edu
Project Information
Center project name: GPVW
Center clone name: CH230-96P15
----- Summary Statistics
Sequencing vector: Plasmid;
Chemistry: Dye-terminator Big Dye 100% of reads
Assembly program: Phrap; version 0.990329
Consensus quality: 116229 bases at least Q40
Consensus quality: 120977 bases at least Q30
Consensus quality: 125402 bases at least Q20
-----
* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html) .
* NOTE: This is a 'working draft' sequence. It currently
* consists of 61 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
1 1076: contig of 1076 bp in length
1077 1176: gap of unknown length
1177 1177: contig of 1129 bp in length
2306 2405: gap of unknown length
2406 3853: contig of 1448 bp in length
3854 3953: gap of unknown length
3954 5443: contig of 1490 bp in length
5444 5543: gap of unknown length
5544 7070: contig of 1527 bp in length
7071 7171: gap of unknown length
7171 8644: contig of 1474 bp in length
8645 8745: gap of unknown length
8745 10281: contig of 1536 bp in length
10281 10380: gap of unknown length
10381 11897: contig of 1517 bp in length
11898 11997: gap of unknown length
11998 13505: contig of 1508 bp in length
13506 13605: gap of unknown length
13606 15171: contig of 1566 bp in length
15172 15271: gap of unknown length
15272 16309: contig of 1038 bp in length
16310 16409: gap of unknown length
16410 17822: contig of 1413 bp in length
17823 17922: gap of unknown length
17923 19009: contig of 1087 bp in length
19010 19109: gap of unknown length
19110 20458: contig of 1349 bp in length
20459 20559: gap of unknown length
20560 22525: contig of 1968 bp in length
22526 22626: gap of unknown length
22627 23925: contig of 1299 bp in length
23926 24025: gap of unknown length
24026 25103: contig of 1078 bp in length
25104 25203: gap of unknown length
25204 26377: contig of 1174 bp in length
26378 26477: gap of unknown length
26478 27960: contig of 1483 bp in length
27961 28060: gap of unknown length
28061 29455: contig of 1395 bp in length

```







```
/replace="t"
variation 1127 /frequency="0.01"
/replace="t"
variation 1130 /frequency="0.01"
/replace="t"
variation 1263 /frequency="0.03"
/replace="a"
variation 1312 /frequency="0.01"
/replace="c"
gene 1440.14478
/gene="IL9R"
join(1440.1646,6769.6882,7312.7423,7540.7718,
8283.8428,9141.9342,9946.10051,11391.11475,
13683.14478)
/product="interleukin 9 receptor"
/gene="IL9R"
variation 1509 /gene="IL9R"
/frequency="0.01"
/replace="a"
join(1619.1646,6769.6882,7312.7423,7540.7718,
8283.8428,9141.9342,9946.10051,11391.11475,
13683.14279)
/gene="IL9R"
/codon_start=1
/product="interleukin 9 receptor"
/protein_id="AAL55435.1"
/db_xref="GI:18071672"
/translation="MGLGRCIWEGWTLSEBALRDMGTLLACICTCVCIGAVSVG
EGQPRSPFTCTNNILRIDCHMSAPELGGSSPWLLFTSNQAPGTHKILGSEC
TVLPPEAVLVPSDNFTTFPHCHMSGRQVSLVDEYIPRRHYKIDPPSDIQNINSSG
HCILTWISIPALIEPTLLSYELAKKQEEWEOQHSDHIVGVTLILAEIDPGF
IHEARLVQMAILEDVVEEBRYTQMSQPCVFOAPORQGPLIPWMPGNTLVA
VISIFLLTGPYTLKLSPRYRIFYONVPSPAMFQPLSYVHNGNFTWMAAGAV
LISODCACTPOGALPEPCVQBATALTGCPAPRWKSVALEEBEGPGRTPGNLSEEDV
LPAGCTEVRVQTLATROEDWAPTSITRPAPDSGRSSSSSSNNNNYCALGCY
GGMHLSPALPGNTQSSGPIPALACGISCHQGLIETQGVAWVLACHQRPGLIHEDLQGM
LIPSVLSKRSMTF"
variation 1727 /gene="IL9R"
/frequency="0.01"
/replace="t"
variation 1801 /gene="IL9R"
/frequency="0.37"
/replace="g"
variation 1809.1810 /gene="IL9R"
/frequency="0.01"
/replace="a"
variation 2059 /gene="IL9R"
/frequency="0.01"
/replace="t"
variation 2316 /gene="IL9R"
/frequency="0.01"
/replace="t"
variation 2584 /gene="IL9R"
/frequency="0.04"
/replace="c"
variation 2606 /gene="IL9R"
/frequency="0.32"
/replace="a"
variation 2676 /gene="IL9R"
/frequency="0.10"
/replace="a"

/replace="t"
variation 2766 /gene="IL9R"
/frequency="0.05"
/replace="a"
variation 2786 /gene="IL9R"
/frequency="0.04"
/replace="t"
variation 2825 /gene="IL9R"
/frequency="0.45"
/replace="a"
variation 2935 /gene="IL9R"
/frequency="0.02"
/replace="a"
variation 3095 /gene="IL9R"
/frequency="0.01"
/replace="a"
variation 3148 /gene="IL9R"
/frequency="0.18"
/replace="g"
variation 3189 /gene="IL9R"
/frequency="0.02"
/replace="c"
variation 3197 /gene="IL9R"
/frequency="0.44"
/replace="a"
variation 3294 /gene="IL9R"
/frequency="0.20"
/replace="g"
variation 3383 /gene="IL9R"
/frequency="0.47"
/replace="c"
variation 3477 /gene="IL9R"
/frequency="0.01"
/replace="a"
variation 3630 /gene="IL9R"
/frequency="0.02"
/replace="c"
variation 3677 /gene="IL9R"
/frequency="0.45"
/replace="t"
repeat_region 3703.4005 /rpt_family="Alu"
/rpt_type="dispersed"
variation 3728 /gene="IL9R"
/frequency="0.18"
/replace="a"
variation 3748 /gene="IL9R"
/frequency="0.01"
/replace="a"
variation 3816 /gene="IL9R"
/frequency="0.30"
/replace="c"
variation 3990 /gene="IL9R"
/frequency="0.30"
/replace="a"
variation 4036 /gene="IL9R"
/frequency="0.10"
/replace="a"
```

```
/gene="IL9R"  
/frequency="0.03"  
/replace="a"
```

Query Match	3.7%	Score 54.8;	DB 9;	Length 15446;
Best Local Similarity	43.7%;	Pred. No. 0.00015;		
Matches 242;	Conservative 0;	Mismatches 312;	Indels 0;	Gaps 0

OY	673	TGTCAATAGTATTCTGAGTGCTGGTAGTCATCGATTGGTGCTCTGCATCACTGAAT	732
Db	12759	TATGTGTCTGTGTGTTCTGTGTGTGTGTCTGTGTGTGTGTGTGTGTGTGTGTGTG	12818
OY	733	CCGGTTGTGTAAATGATGCTCAGACCCTTGATAAACAACGTGTGGATGACTGTGTGAC	792
Db	12819	TGTGTGTGTATTATGTGTGTGTGTCTGTGTGTGTGTGTAAAGTCTGTGTGTGTGTGTG	12878
OY	793	AATTACAGTAAGCAGAGTGAATTTCCATCTCCGTCAATGTGTGTGGTAGGTGAGCTGA	852
Db	12879	TGTCTCTGTGTGTGTCTGTATCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTG	12938
OY	853	TGTGAGGTGAGTTAGAGTATGTGATGGATGGAATGTGATGTATGGAGATGATGAGA	912
Db	12939	TCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTG	12998
OY	913	CCAATTGAGAGAGAGAGAACATCATGAATATCTATGCAAAACGATGAGAGTGTGTGTA	972
Db	12999	TGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTC	13058
OY	973	CATGAAGTAAAGGGGCTCATGTGTGTGTGAGTATGTGTGTGTGGAGAGTGTGTGATGAGAGTGC	1032
Db	13059	TGTGTGTGTATGTGTCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTG	13118
OY	1033	ACTTATGATGCAATATATAGTGCATGTGTATTTGTATGTGTGTGTGTGTGTGTGTGTGTG	1092
Db	13119	CTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTCT	13178
OY	1093	GTAAGAAGGTGATATTGTGCCAGTTGATATTATTCCAAATTAACCTTGTTGTGTGTG	1152
Db	13179	GTTGTGTGTATTATGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTG	13238
OY	1153	TTCGGTTGTGTGTGTGTCTTCTTCTCATCTGTACTGTACGTGTGTGTGTGTGTG	1212
Db	13239	TATCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTG	13298
OY	1213	TGGTTGTTGTGTG	1226
Db	13299	TGTGTGTGTGTGTG	13312

RESULT	37
AC120787/c	
LOCUS	AC120787
DEFINITION	Mus musculus clone RP24-180L12, LOW-PASS SEQUENCE SAMPLING.
ACCESSION	AC120787
VERSION	AC120787.1 GI:20514883
KEYWORDS	HTG; HTGS PHASE0.
SOURCE	Mus musculus.
ORGANISM	Mus musculus.
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus. 1 (bases 1 to 65472)
AUTHORS	Birtren,B., Linton,L., Nusbaum,C. and Lander,E.
TITLE	Mus musculus, clone RP24-180L12
JOURNAL	unpublished
REFERENCE	2 (bases 1 to 65472)
AUTHORS	Birtren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N.,

Bouknightalter, B., Brown, A., Canarata, J., Campodiano, A., Chang, J., Chazaro, S., Chepel, Y., Colangelo, M., Collins, S., Collimore, A., Cook, A., Cooke, P., Deatrellano, K., Dewar, K., Diaz, J.S., Dodge, S., Fero, S., Ferreira, P., FitzHugh, W., Gage, D., Galagan, J., Gardyna, S., Gird, S., Goyette, M., Graham, L., Grand-Pierre, N., Hages, B., Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C.,

TITLE  
JOURNAL  
COMMENT

\* NOTE: This record contains 82 individual  
\* sequencing reads that have not been assembled into  
\* contigs. Runs of N are used to separate the reads  
\* and the order in which they appear is completely  
\* arbitrary. Low-pass sequence sampling is useful for  
\* identifying clones that may be gene-rich and allows  
\* overlap relationships among clones to be deduced.  
\* However, it should not be assumed that this clone  
\* will be sequenced to completion. In the event that  
\* the record is updated, the accession number will  
\* be preserved.

1 673: contig of 673 bp in length

674 773: gap of 100 bp

774 1466: contig of 693 bp in length

1467 1566: gap of 100 bp

1567 2254: contig of 688 bp in length

2255 2354: gap of 100 bp

2355 3063: contig of 709 bp in length

3064 3163: gap of 100 bp

3164 3873: contig of 710 bp in length

3874 3973: gap of 100 bp

3974 4648: contig of 675 bp in length

4649 4748: gap of 100 bp

4749 5431: contig of 683 bp in length

5432 5531: gap of 100 bp

5532 6228: contig of 697 bp in length

6229 6328: gap of 100 bp

6329 7029: contig of 701 bp in length

7030 7129: gap of 100 bp

7130 7819: contig of 690 bp in length

7820 7919: gap of 100 bp

7920 8620: contig of 701 bp in length

8621 8720: gap of 100 bp

8721 9404: contig of 684 bp in length

9405 9504: gap of 100 bp

9505 10211: contig of 707 bp in length

10212 10311: gap of 100 bp

10312 11007: contig of 656 bp in length

11008 11107: gap of 100 bp

11108 11811: contig of 704 bp in length

11812 11911: gap of 100 bp

11912 12613: contig of 702 bp in length

12614 12713: gap of 100 bp

12714 13351: contig of 638 bp in length

Center project name: 180\_L\_12

Center clone name: 180\_L\_12

NOTE: This record contains 82 individual  
\* sequencing reads that have not been assembled into  
\* contigs. Runs of N are used to separate the reads  
\* and the order in which they appear is completely  
\* arbitrary. Low-pass sequence sampling is useful for  
\* identifying clones that may be gene-rich and allows  
\* overlap relationships among clones to be deduced.  
\* However, it should not be assumed that this clone  
\* will be sequenced to completion. In the event that  
\* the record is updated, the accession number will  
\* be preserved.

1 673: contig of 673 bp in length

674 773: gap of 100 bp

774 1466: contig of 693 bp in length

1467 1566: gap of 100 bp

1567 2254: contig of 688 bp in length

2255 2354: gap of 100 bp

2355 3063: contig of 709 bp in length

3064 3163: gap of 100 bp

3164 3873: contig of 710 bp in length

3874 3973: gap of 100 bp

3974 4648: contig of 675 bp in length

4649 4748: gap of 100 bp

4749 5431: contig of 683 bp in length

5432 5531: gap of 100 bp

5532 6228: contig of 697 bp in length

6229 6328: gap of 100 bp

6329 7029: contig of 701 bp in length

7030 7129: gap of 100 bp

7130 7819: contig of 690 bp in length

7820 7919: gap of 100 bp

7920 8620: contig of 701 bp in length

8621 8720: gap of 100 bp

8721 9404: contig of 684 bp in length

9405 9504: gap of 100 bp

9505 10211: contig of 707 bp in length

10212 10311: gap of 100 bp

10312 11007: contig of 656 bp in length

11008 11107: gap of 100 bp

11108 11811: contig of 704 bp in length

11812 11911: gap of 100 bp

11912 12613: contig of 702 bp in length

12614 12713: gap of 100 bp

12714 13351: contig of 638 bp in length

Center project name: 180\_L\_12

Center clone name: 180\_L\_12

NOTE: This record contains 82 individual  
\* sequencing reads that have not been assembled into  
\* contigs. Runs of N are used to separate the reads  
\* and the order in which they appear is completely  
\* arbitrary. Low-pass sequence sampling is useful for  
\* identifying clones that may be gene-rich and allows  
\* overlap relationships among clones to be deduced.  
\* However, it should not be assumed that this clone  
\* will be sequenced to completion. In the event that  
\* the record is updated, the accession number will  
\* be preserved.

1 673: contig of 673 bp in length

674 773: gap of 100 bp

774 1466: contig of 693 bp in length

1467 1566: gap of 100 bp

1567 2254: contig of 688 bp in length

2255 2354: gap of 100 bp

2355 3063: contig of 709 bp in length

3064 3163: gap of 100 bp

3164 3873: contig of 710 bp in length

3874 3973: gap of 100 bp

3974 4648: contig of 675 bp in length

4649 4748: gap of 100 bp

4749 5431: contig of 683 bp in length

5432 5531: gap of 100 bp

5532 6228: contig of 697 bp in length

6229 6328: gap of 100 bp

6329 7029: contig of 701 bp in length

7030 7129: gap of 100 bp

7130 7819: contig of 690 bp in length

7820 7919: gap of 100 bp

7920 8620: contig of 701 bp in length

8621 8720: gap of 100 bp

8721 9404: contig of 684 bp in length

9405 9504: gap of 100 bp

9505 10211: contig of 707 bp in length

10212 10311: gap of 100 bp

10312 11007: contig of 656 bp in length

11008 11107: gap of 100 bp

11108 11811: contig of 704 bp in length

11812 11911: gap of 100 bp

11912 12613: contig of 702 bp in length

12614 12713: gap of 100 bp

12714 13351: contig of 638 bp in length

Center project name: 180\_L\_12

Center clone name: 180\_L\_12

NOTE: This record contains 82 individual  
\* sequencing reads that have not been assembled into  
\* contigs. Runs of N are used to separate the reads  
\* and the order in which they appear is completely  
\* arbitrary. Low-pass sequence sampling is useful for  
\* identifying clones that may be gene-rich and allows  
\* overlap relationships among clones to be deduced.  
\* However, it should not be assumed that this clone  
\* will be sequenced to completion. In the event that  
\* the record is updated, the accession number will  
\* be preserved.

1 673: contig of 673 bp in length

674 773: gap of 100 bp

774 1466: contig of 693 bp in length

1467 1566: gap of 100 bp

1567 2254: contig of 688 bp in length

2255 2354: gap of 100 bp

2355 3063: contig of 709 bp in length

3064 3163: gap of 100 bp

3164 3873: contig of 710 bp in length

3874 3973: gap of 100 bp

3974 4648: contig of 675 bp in length

4649 4748: gap of 100 bp

4749 5431: contig of 683 bp in length

5432 5531: gap of 100 bp

5532 6228: contig of 697 bp in length

6229 6328: gap of 100 bp

6329 7029: contig of 701 bp in length

7030 7129: gap of 100 bp

7130 7819: contig of 690 bp in length

7820 7919: gap of 100 bp

7920 8620: contig of 701 bp in length

8621 8720: gap of 100 bp

8721 9404: contig of 684 bp in length

9405 9504: gap of 100 bp

9505 10211: contig of 707 bp in length

10212 10311: gap of 100 bp

10312 11007: contig of 656 bp in length

11008 11107: gap of 100 bp

11108 11811: contig of 704 bp in length

11812 11911: gap of 100 bp

11912 12613: contig of 702 bp in length

12614 12713: gap of 100 bp

12714 13351: contig of 638 bp in length

Center project name: 180\_L\_12

Center clone name: 180\_L\_12

NOTE: This record contains 82 individual  
\* sequencing reads that have not been assembled into  
\* contigs. Runs of N are used to separate the reads  
\* and the order in which they appear is completely  
\* arbitrary. Low-pass sequence sampling is useful for  
\* identifying clones that may be gene-rich and allows  
\* overlap relationships among clones to be deduced.  
\* However, it should not be assumed that this clone  
\* will be sequenced to completion. In the event that  
\* the record is updated, the accession number will  
\* be preserved.

1 673: contig of 673 bp in length

674 773: gap of 100 bp

774 1466: contig of 693 bp in length

1467 1566: gap of 100 bp

1567 2254: contig of 688 bp in length

2255 2354: gap of 100 bp

2355 3063: contig of 70



REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
REFERENCE  
AUTHORS

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
1 (bases 1 to 156238)  
Biren, B., Linton, L., Nusbaum, C. and Lander, E.  
Mus musculus, clone RP23-12419  
Unpublished  
2 (bases 1 to 156238)

REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
REFERENCE  
AUTHORS

1 (bases 1 to 156238)  
Biren, B., Linton, L., Nusbaum, C., Lander, E., All, A., Allen, N.,  
Anderson, S., Barina, N., Bastien, V., Boguslavsky, L., Bokhgalter, B.,  
Brown, A., Camarata, J., Campopiano, A., Chang, J., Chazaro, B.,  
Chapel, Y., Colangelo, M., Collins, S., Collamore, A., Cook, A.,  
Cooke, P., Dearellano, K., Dewar, K., Diaz, J.S., Dodge, S., Faro, S.,  
Ferreira, P., Fitzhugh, W., Gage, D., Galagan, J., Gately, S.,  
Ginde, S., Gord, S., Goyette, M., Graham, L., Grand-Pierre, N.,  
Hagos, B., Heatford, A., Horton, L., Hulme, W., Iliev, I., Johnson, R.,  
Jones, C., Kamat, A., Karatas, A., Kells, C., LaRoque, K.,  
Lamazares, R., Landers, T., Lehoczy, J., Levine, R., Liu, G.,  
Maclean, C., Macdonald, P., Major, U., Marguis, N., Matthews, C.,  
McCarthy, M., McEwan, P., McKernan, K., McPheeters, R., Meldrum, J.,  
Meneus, L., Mihova, T., Mienga, V., Murphy, T., Naylor, J., Nguyen, C.,  
Norbu, C., Norman, C.H., O'Connor, T., O'Donnell, P., O'Neill, D.,  
Oliver, J., Peterson, K., Phunkhang, P., Pierre, N., Pollara, V.,  
Raymond, C., Roesti, M., Riley, R., Rise, C., Rogov, P.,  
Roman, J., Roesti, M., Roy, A., Santos, R., Schauer, S., Schupbach, R.,  
Seaman, S., Severy, P., Spencer, B., Stange-Thomann, N., Stojanovic, N.,  
Strauss, N., Subramanian, A., Talamas, J., Tesfaye, S., Theodore, J.,  
Topham, K., Travers, M., Travis, N., Triggillo, J., Vassiliev, H.,  
Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W.J., Young, G.,  
Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.

REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
REFERENCE  
AUTHORS

Direct Submission  
Submitted (22-DEC-2001) Whitehead Institute/MIT Center for Genome  
Research, 320 Charles Street, Cambridge, MA 02141, USA  
3 (bases 1 to 156238)

REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
REFERENCE  
AUTHORS

1 (bases 1 to 156238)  
Biren, B., Linton, L., Nusbaum, C., Lander, E., All, A., Allen, N.,  
Anderson, S., Barina, N., Bastien, V., Bloom, T., Boguslavsky, L.,  
Bokhgalter, B., Brown, A., Camarata, J., Campopiano, A., Chang, J.,  
Chazaro, B., Chapel, Y., Colangelo, M., Collins, S., Collamore, A.,  
Cook, A., Cooke, P., Dearellano, K., Dewar, K., Diaz, J.S., Dodge, S.,  
Faro, S., Ferreira, P., Fitzhugh, W., Gage, D., Galagan, J., Gately, S.,  
Ginde, S., Gord, S., Goyette, M., Graham, L., Grand-Pierre, N.,  
Hagos, B., Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C.,  
Kamat, A., Karatas, A., Kells, C., LaRoque, K., Lamazares, R.,  
Landers, T., Lehoczy, J., Levine, R., Lindblad-Toh, K., Liu, G.,  
Maclean, C., Macdonald, P., Major, U., Marguis, N., Matthews, C.,  
McCarthy, M., McEwan, P., McKernan, K., Meldrum, J., Meneus, L.,  
Mihova, T., Mienga, V., Murphy, T., Naylor, J., Nguyen, C., Nicol, R.,  
Norbu, C., Norman, C.H., O'Connor, T., O'Donnell, P., O'Neill, D.,  
Oliver, J., Peterson, K., Phunkhang, P., Pierre, N., Pollara, V.,  
Raymond, C., Roesti, M., Riley, R., Rise, C., Rogov, P.,  
Roman, J., Roesti, M., Roy, A., Santos, R., Schauer, S., Schupbach, R.,  
Seaman, S., Severy, P., Spencer, B., Stange-Thomann, N., Stojanovic, N.,  
Strauss, N., Subramanian, A., Talamas, J., Tesfaye, S., Theodore, J.,  
Topham, K., Travers, M., Travis, N., Triggillo, J., Vassiliev, H.,  
Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W.J., Young, G.,  
Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.

REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
REFERENCE  
AUTHORS

Direct Submission  
Submitted (05-APR-2002) Whitehead Institute/MIT Center for Genome  
Research, 320 Charles Street, Cambridge, MA 02141, USA  
All repeats were identified using RepeatMasker:  
Smt, A.F.A. & Green, P. (1996-1997)  
http://ftp.genome.washington.edu/RM/RepeatMasker.html

REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
REFERENCE  
AUTHORS

Center: Whitehead Institute/ MIT Center for Genome Research  
Center code: WIRB  
Web site: http://www-seq.wi.mit.edu  
Contact: sequence\_submissions@genome.wi.mit.edu  
Project Information  
Center project name: I15094  
Center clone name: 124.1.9  
Summary Statistics  
Sequencing vector: Plasmid; n/a; 100% of reads  
Chemistry: Dye-terminator Big Dye; 100% of reads  
Assembly program: Phrap; version 0.960731

Consensus quality: 150640 bases at least Q40  
Consensus quality: 152759 bases at least Q30  
Consensus quality: 153421 bases at least Q20  
Insert size: 168000; agarose-fp  
Insert size: 154038; sum-of-contigs  
Quality coverage: 7.6 in Q20 bases; agarose-fp  
Quality coverage: 8.3 in Q20 bases; sum-of-contigs

NOTE: This is a 'working draft' sequence. It currently  
consists of 23 contigs. Gaps between the contigs  
are represented as runs of N. The order of the pieces  
is believed to be correct as given, however the sizes  
of the gaps between them are based on estimates that have  
been provided by the submitter.

This sequence will be replaced  
by the finished sequence as soon as it is available and  
the accession number will be preserved.

1 9659: contig of 9659 bp in length  
9660 9759: gap of 100 bp  
9760 10423: contig of 664 bp in length  
10424 10523: gap of 100 bp  
10524 11140: contig of 617 bp in length  
11141 11240: gap of 100 bp  
11241 12084: contig of 844 bp in length  
12085 12184: gap of 100 bp  
12185 13496: contig of 1312 bp in length  
13497 13596: gap of 100 bp  
13597 15640: contig of 2044 bp in length  
15641 15740: gap of 100 bp  
15741 17956: contig of 2256 bp in length  
17957 18096: gap of 100 bp  
18097 19704: contig of 1608 bp in length  
19705 19804: gap of 100 bp  
19805 22393: contig of 2589 bp in length  
22394 22493: gap of 100 bp  
22494 24061: contig of 1568 bp in length  
24062 24161: gap of 100 bp  
24162 26022: contig of 1861 bp in length  
26023 26123: gap of 100 bp  
26123 29016: contig of 2894 bp in length  
29017 29116: gap of 100 bp  
29117 32239: contig of 3123 bp in length  
32240 32339: gap of 100 bp  
32340 36183: contig of 3844 bp in length  
36184 36283: gap of 100 bp  
36284 38504: contig of 2221 bp in length  
38505 38604: gap of 100 bp  
38605 42504: contig of 3900 bp in length  
42505 42604: gap of 100 bp  
42605 50019: contig of 7415 bp in length  
50020 50119: gap of 100 bp  
50120 96712: contig of 48593 bp in length  
96713 96812: gap of 100 bp  
96813 11186: contig of 12374 bp in length  
11187 11286: gap of 100 bp  
11287 120573: contig of 9287 bp in length  
120574 120673: gap of 100 bp  
120674 133299: contig of 12626 bp in length  
133300 133399: gap of 100 bp  
133400 151103: contig of 17704 bp in length  
151104 151203: gap of 100 bp  
151204 156238: contig of 5035 bp in length.

# FEATURES

Source  
1. 156238  
/organism="Mus musculus"  
/db\_xref="taxon:10090"  
/clone="RP23-12419"  
/clone\_1lb="RP23-12419"  
/note="assembly\_fragment  
clone\_end:SP6  
vector\_side:left"  
misc\_feature  
9760..10423

```
misc_feature /note="assembly_fragment" 10524. 11140
              /note="assembly_fragment"
misc_feature 11241. 12084
              /note="assembly_fragment"
misc_feature 12185. 13496
              /note="assembly_fragment"
misc_feature 13597. 15640
              /note="assembly_fragment"
misc_feature 15741. 17996
              /note="assembly_fragment"
misc_feature 18097. 19704
              /note="assembly_fragment"
misc_feature 19805. 22393
              /note="assembly_fragment"
misc_feature 22494. 24061
              /note="assembly_fragment"
misc_feature 24162. 26022
              /note="assembly_fragment"
misc_feature 26123. 29016
              /note="assembly_fragment"
misc_feature 29117. 32239
              /note="assembly_fragment"
misc_feature 32340. 36183
              /note="assembly_fragment"
misc_feature 36284. 38504
              /note="assembly_fragment"
misc_feature 38605. 42504
              /note="assembly_fragment"
misc_feature 42605. 50019
              /note="assembly_fragment"
misc_feature 50120. 98712
              /note="assembly_fragment"
misc_feature 98813. 11186
              /note="assembly_fragment"
misc_feature 111287. 120573
              /note="assembly_fragment"
misc_feature 120674. 133299
              /note="assembly_fragment"
misc_feature 133400. 151103
              /note="assembly_fragment"
misc_feature 151204. 156238
              /note="assembly_fragment"
clone_end:T7
vector_side::right"
```

Query Match	3.7%;	Score 54.8;	DB 2;	Length 156238;
Best Local Similarity	47.4%;	Pred. No. 0.00018;		
Matches 164;	Conservative 0;	Mismatches 182;	Indels 0;	Gaps 0

[illegible]

RESULT_39			
LOCUS	HSA271736	158661 bp	DNA linear PRI 08-JUL-2000
DEFINITION	Homo sapiens Xq pseudautosomal region; segment 2/2.		
ACCESSION	AJ271736.1	GI:8979791	
VERSION	CX00P1.1	ILR gene; interleukin 9 receptor; repeat region;	
KEYWORDS	repetitive DNA; SYBL1 gene; synapobrevin-like 1 protein.		
SOURCE	human.		
ORGANISM	Homo sapiens		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.		
AUTHORS	Migliaccolo,A., D'Esposito,M., Esposito,T., Gianfrancesco,F., Cuccia,S., Meradante,G., Pamone,E., Archidiacono,N., Rocchi,M., Schlessinger,D. and D'Urso,M. Differentially regulated and evolved genes in the fully sequenced Xq/Yq pseudautosomal region Hum. Mol. Genet. 9 (3), 395-401 (2000)		
TITLE	JOURNAL MEDLINE 20122249 10655549 2 (bases 1 to 158661) Ciccodicola,A. Direct Submission Genetics and Biophysics (IIGB-CNR), Via P. Castellino 111, Napoli, 80131, ITALY		
FEATURES	source		
	Location/Qualifiers		
	1..158661		
	/organism="Homo sapiens"		
	/db_xref="taxon:9606"		
	/chromosome="X"		
	/map="Xq28"		
	801..867		
	/note="LINE2B d"		
	1566..1647		
	/note="LINE2 d"		
	2904..3018		
	/note="MIR d"		
	3046..3415		
	/note="LINE2 c"		
	3810..3939		
	/note="LIPB3 c"		
	3943..4165		
	/note="LIMA2 d"		
	4259..4318		
	/note="LIMA2 d"		
	5824..5982		
	/note="Alu Jo d"		
	6384..6687		
	/note="LIPAL1 d"		
	6851..6902		
	/note="MIR c"		
	7213..7436		
	/note="MER41B c"		
	7437..7506		
	/note="MSTA c"		
	7870..7959		
	/note="LINE2 c"		
	7963..8251		
	/note="Alu Sg d"		
	8267..8595		
	/note="LINE2 c"		
	9144..71522		
	/gene="SYBL1"		
	join(9144..9213,17222..17376,23387..23444,25877..26014,		
	28262..28352,47578..47645,67466..67558,69648..71522)		
	/gene="SYBL1"		
	9144..9213		
exon			

```
/gene="SYBL1"
/number=1
repeat_region 9474..9610
/note="LINE2 d"
repeat_region 9961..10098
/note="LINE2 c"
repeat_region 10548..10628
/note="LINE2B c"
repeat_region 11485..11775
/note="Alu Sx c"
repeat_region 11901..12190
/note="Alu Sp c"
repeat_region 13245..13726
/note="MLT1C d"
repeat_region 14209..14474
/note="L3 c"
repeat_region 14549..14676
/note="CRI_HS c"
repeat_region 15077..15385
/note="LINE2 c"
repeat_region 15411..15709
/note="Alu Jb d"
repeat_region 16075..16225
/note="MER5A d"
repeat_region 16536..16775
/note="MIR d"
repeat_region 16972..16993
/note="(GT) d"
exon 17222..17376
/gene="SYBL1"
/number=2
join(17231..17376,23387..23444,25877..26014,28262..28352,
47578..47645,67466..67558,69648..69716)
/gene="SYBL1"
/codon_start=1
/product="synaptoobrevin-like 1 protein"
/protein_id="CAB96816.1"
/db_xref="GI:8979792"
/translation="MALIPAVARGTTTAAKMGCGNLEVTQIILAKIPSENNKLT
YSHGNYLFHYICQDRIVYLCTDDPFERSRAFNFLNEIKKRPQTTGSRACATLAFYAM
NSEFSSVLAQLKHSENNKGLDKWETQAOVDELGIWRNIDLVAGGERLELIDK
TENLVDSVTFKTSRNILARAMCMKMLKLTIIIVSIYIIVSPICGGFTWPSCV
KK"
repeat_region 17687..17811
/note="LINE2B c"
repeat_region 17858..18112
/note="LINE2 c"
repeat_region 18115..18134
/note="(TTTA) d"
repeat_region 18135..18424
/note="Alu Sx c"
repeat_region 18880..19037
/note="MLT1J c"
repeat_region 19759..19867
/note="MLT1J1 c"
repeat_region 20116..20333
/note="LINE2 d"
repeat_region 20506..20525
/note="(TTA) d"
repeat_region 20526..20816
/note="Alu Sq c"
repeat_region 20943..21247
/note="LINE2 c"
repeat_region 21252..21592
/note="MLT1A c"
repeat_region 21599..21746
/note="LINE2 c"
repeat_region 21975..22366
/note="LINE2 d"
repeat_region 22062..22086
/note="(T) d"
repeat_region 22392..22683
/note="Alu Jb c"
```

```
repeat_region 22695..22773
/note="LINE2B d"
repeat_region 22887..22976
/note="LINE2 d"
exon 23387..23444
/gene="SYBL1"
/number=3
repeat_region 23518..23852
/note="MER1B d"
repeat_region 24483..24571
/note="MLT1L c"
repeat_region 24663..25074
/note="MER80 d"
repeat_region 25641..25802
/note="MIR d"
exon 25877..26014
/gene="SYBL1"
/number=4
repeat_region 26228..26295
/note="L3 c"
repeat_region 26419..26637
/note="LINE2 c"
repeat_region 26638..27136
/note="LINE2 c"
repeat_region 27202..27408
/note="X55"
exon 28262..28352
/gene="SYBL1"
/number=5
exon 28353..47577
/gene="SYBL1"
/number=5
repeat_region 29286..30995
/note="L1 c"
repeat_region 30996..31230
/note="L1 d"
repeat_region 31233..32121
/note="L1PA2 d"
repeat_region 32199..33037
/note="LINE7 c"
repeat_region 32653..32681
/note="(AT) d"
repeat_region 33864..34015
/note="MER31B d"
repeat_region 34038..34290
/note="MER31B d"
repeat_region 35046..35717
/note="SVA c"
repeat_region 35718..35742
/note="SVA c"
repeat_region 35743..35912
/note="SVA c"
repeat_region 35913..36030
/note="Alu Sbcg d"
repeat_region 36056..36293
/note="Alu Sz d"
repeat_region 36294..36339
/note="SVA c"
```

Query Match 3.7%; Score 54.8; DB 9; length 158661;  
Best Local Similarity 43.7%; Pred.No. 0.00018;  
Matches 242; Conservative 0; Mismatches 312; Indels 0; Gaps 0;

```
QY 673 TGTCACTCAGTATTCTGTGTCTGTGTATCTGTTTGTGTCTGTCTGTACTGAT 732
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 136658 TATGTGTCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTG 136717
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 733 CCGATTGTGTATTGTGTGAGACCTTGATATACACAGCGTGTGATGACTTGTGAC 792
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 136718 TGTGTGTGTATTATGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTG 136777
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 793 AGTTAGTAGCAGAGTGAATTCATCTCGGTCATTGTGTGTGTGAGGTGACGTGA 852
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
```

Oquague,N., Ovidio,R., Pace,A., Payton,B., Peery,J., Perez,L.,  
 Peters,L., Pickens,R., Primus,E., Pu,L.L., Quiles,M., Ren,Y.,  
 Rives,M., Rojcs,A., Rojnbokan,I., Rolfe,M., Ruiz,S., Savery,G.,  
 Scherer,S., Scott,G., Shen,H., Shoochari,N., Sisson,I.,  
 Sodejgren,E., Sonaikc,T., Sparks,A., Stanley,H., Stone,H.,  
 Sutton,A., Svatek,A., Tabor,P., Tamerisa,A., Tamerisa,K., Tang,H.,  
 Tansey,J., Taylor,C., Taylor,T., Telford,B., Thomas,N., Thomas,S.,  
 Usmani,K., Vasquez,L., Vera,V., Villalon,D., Vinson,R., Wang,Q.,  
 Wang,S., Ward-Moore,S., Warren,R., Washington,C., Wellington,S.,  
 Williams,G., Williamson,A., Wiczeky,R., Wooden,S., Worley,K.,  
 Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D.,  
 Weinstein,G. and Gibbs,R.  
 Direct Submission  
 Unpublished  
 2 (bases 1 to 176469)  
 Worley,K.C.  
 Direct Submission  
 Submitted (15-MAY-2002) Human Genome Sequencing Center, Department  
 of Molecular and Human Genetics, Baylor College of Medicine, One  
 Baylor Plaza, Houston, TX 77030, USA  
 3 (bases 1 to 176469)  
 Worley,K.C.  
 Direct Submission  
 Submitted (23-JUL-2002) Human Genome Sequencing Center, Department  
 of Molecular and Human Genetics, Baylor College of Medicine, One  
 Baylor Plaza, Houston, TX 77030, USA  
 On Jul 18, 2002 this sequence version replaced gi:20799981.  
 ----- Genome Center  
 Center: Baylor College of Medicine

RESULT	LOCUS	DEFINITION	ACCESSION	VERSION	KEYWORDS	SOURCE	ORGANISM
40	ACT121034	Rattus norvegicus clone CH230-43C5, *** SEQUENCING IN PROGRESS ***.	176469 bp	DNA	linear	HTG 23-JUL-2007	
		68 unordered pieces.					
	ACT121034						
	ACT121034.2	G1:21903217					
	HTG; HTGS PHASE1.						
	Rattus norvegicus.						
	Rattus norvegicus						
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;						
	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;						
	Rattus.						
	1 (bases 1 to 176469)						
	Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C.,						
	Alsbrooks,S.L., Amarantunga,H.C., Are,J.R., Ayele,M., Banks,T.,						
	Babbaria,J., Benton,J., Blinage,K., Blankenburg,K., Bonnin,D.,						
	Bouchak,J., Bowls,P., Brieva,M., Brown,E., Brown,M., Bryant,N.P.,						
	Buhay,C., Burch,P., Burkett,C., Butrell,K.L., Byrd,N.C.,						
	Chen,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D.,						
	Clegg,G., Chen,R., Chen,Z., Chowdhury,I., Christopoulos,C.,						
	Cleveland,L.C.D., Cox,C., Coyle,M.D., Dathornes,S.R., David,R.,						
	Dayla,M.D., Davis,C., Davy-Carroll,L., Dedrich,D.A.,						
	DeJarey,K.R., Delgado,O., Demn,A.L., Ding,Y., Dinl,H.R.,						
	Douthwaite,K.J., Draper,H., Duncan-Rocha,S., Durbin,K.J.,						
	Earnhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Escotto,M.,						
	Falls,T., Ferraguto,D., Flisgyn,N., Ford,T., Foster,P., Frantz,P.,						
	Gabisi,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R.,						
	Goirell,J.H., Guevara,W., Gunaratne,P., Hale,S., Hamilton,K.,						
	Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A., Hernandez,J.,						
	Hernandez,O., Hodgson,A., Hogues,N., Holloway,C., Hollins,B.E.,						
	Homesi,F., Howard,S., Huber,J., Hulik,S., Hume,J., Jackson,L.E.,						
	Jacobson,B., Jia,Y., Johnson,R., Jollivet,S., Joudan,S.,						
	Karlsson,E., Kelly,S., Khan,U., King,L., Korvan,J., Kovar,C.,						
	Kracovic,U., Kureshi,A., Landry,N., Leal,B., Lewis,L.C., Lewis,L.,						
	Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W., Louisedge,H.,						
	Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R., Ma,J.,						
	Maneshwari,M., Mapua,P., Martin,R., Martindale,A., Martinez,B.,						
	Masey-Eby,Machiney,E., McLeod,M.P., Meador,M., Mei,G., Metzler,M.,						
	Minter,G., Minter,Z., Mitchell,T., Mohabbat,K., Morgan,M., Morris,S.,						
	Moser,M., Neal,D., Newton,J., Newton,S., Nguyen,A., Nguyen,N.,						
	Nguyen,N., Nickerson,E., Nwokhekwo,S., Ogun,M., Okunolu,G.,						

```

Contact: hgsc-help@bcm.tmc.edu
----- Project Information -----
Center project name: CH230-43CS
Center clone name: CH230-43CS
Summary Statistics
Sequencing vector: Plasmid;
Chemistry: Dye-terminator Big Dye: 100% of reads
Assembly program: Phrap; version 0.990329
Consensus quality: 120116 bases at least Q40
Consensus quality: 125734 bases at least Q30
Consensus quality: 128936 bases at least Q20
-----
* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/genbank\_draft\_data.html)
* NOTE: This is a 'working draft' sequence. It currently
* consists of 68 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
* 1
* 1165: contig of 1165 bp in length
* 1166
* 1265: gap of unknown length
* 1275: contig of 1010 bp in length
* 2276
* 2375: gap of unknown length
* 3587: contig of 1212 bp in length
* 3588
* 3687: gap of unknown length
* 3688
* 5389: contig of 1702 bp in length
* 5390
* 5489: gap of unknown length
* 5490
* 6871: contig of 1382 bp in length
* 6872
* 6971: gap of unknown length
* 6972
* 8131: contig of 1160 bp in length
* 8132
* 8231: gap of unknown length
* 8232
* 9941: contig of 1710 bp in length
* 9942
* 10041: gap of unknown length
* 10042
* 11153: contig of 1112 bp in length
* 11154
* 11253: gap of unknown length
* 11254
* 12546: contig of 1293 bp in length
* 12547
* 12646: gap of unknown length
* 12647
* 13646: contig of 1000 bp in length
* 13647
* 13746: gap of unknown length

```



*	13747	15146	contig of 1600 bp in length
*	15347	15446	gap of unknown length
*	15447	16703	contig of 1257 bp in length
*	16704	16803	gap of unknown length
*	16804	16842	contig of 1739 bp in length
*	18543	18642	gap of unknown length
*	18643	19707	contig of 1065 bp in length
*	19708	19907	gap of unknown length
*	19808	21127	contig of 1920 bp in length
*	21768	21827	gap of unknown length
*	21828	23306	contig of 1479 bp in length
*	23307	23406	gap of unknown length
*	23407	25398	contig of 1992 bp in length
*	25399	25498	gap of unknown length
*	25499	27431	contig of 1933 bp in length
*	27432	27531	gap of unknown length
*	27532	28842	contig of 1311 bp in length
*	28843	28942	gap of unknown length
*	28943	30257	contig of 1315 bp in length
*	30258	30357	gap of unknown length
*	30358	31719	contig of 1362 bp in length
*	31720	31819	gap of unknown length
*	31820	33184	contig of 1365 bp in length
*	33185	33284	gap of unknown length
*	33285	34320	contig of 1036 bp in length
*	34321	34420	gap of unknown length
*	34421	36489	contig of 2069 bp in length
*	36490	36589	gap of unknown length
*	36590	39126	contig of 2537 bp in length
*	39127	39226	gap of unknown length
*	39227	40323	contig of 1207 bp in length
*	40343	40533	gap of unknown length
*	40534	41644	contig of 1111 bp in length
*	40534	41644	gap of unknown length
*	41645	41744	gap of unknown length
*	41745	43746	contig of 2002 bp in length
*	43747	43846	gap of unknown length
*	43847	45909	contig of 2063 bp in length
*	45910	46009	gap of unknown length
*	46010	47913	contig of 1904 bp in length
*	47914	48013	gap of unknown length
*	48014	49774	contig of 1761 bp in length
*	49775	49874	gap of unknown length
*	49875	51187	contig of 1313 bp in length
*	51188	51287	gap of unknown length
*	51288	53387	contig of 2100 bp in length
*	53387	53487	gap of unknown length
*	53488	55530	contig of 2043 bp in length
*	55531	55630	gap of unknown length
*	55631	55693	contig of 1063 bp in length
*	56794	56793	gap of unknown length
*	56794	58106	contig of 1313 bp in length
*	58107	58206	gap of unknown length
*	58207	60545	contig of 2339 bp in length
*	60546	60645	gap of unknown length
*	60646	63686	contig of 3041 bp in length
*	63687	63786	gap of unknown length
*	63787	66870	contig of 3084 bp in length
*	66871	66970	gap of unknown length
*	66971	66881	contig of 1611 bp in length
*	66882	66881	gap of unknown length
*	66882	70295	contig of 1614 bp in length
*	70296	70395	gap of unknown length
*	70396	72420	contig of 2025 bp in length
*	72421	72520	gap of unknown length
*	72521	74958	contig of 2438 bp in length
*	74959	75058	gap of unknown length
*	75059	76706	contig of 1648 bp in length
*	76707	76806	gap of unknown length
*	76807	78176	contig of 1370 bp in length
*	78177	78276	gap of unknown length
*	78277	80729	contig of 2453 bp in length
*	80730	80829	gap of unknown length
*	80830	83988	contig of 3159 bp in length

*	83989	84088:	gap of unknown length
*	84089	86241:	contig of 2153 bp in length
*	86242	86341:	gap of unknown length
*	89330:	contig of 2889 bp in length	
*	89321	89330:	gap of unknown length
*	89331	91850:	contig of 2520 bp in length
*	91851	91950:	gap of unknown length
*	91951	94310:	contig of 2360 bp in length
*	94311	94410:	gap of unknown length
*	94411	98405:	contig of 3995 bp in length
*	98406	98505:	gap of unknown length
*	98506	101050:	contig of 2545 bp in length
*	101051	101150:	gap of unknown length
*	101151	103868:	contig of 2718 bp in length

Query Match	3.7%	Score 54.8;	DB 2;	Length 176469;
Best Local Similarity	51.6%	Pred. No. 0.00018;		
Matches 113; Conservative	0;	Mismatches 106;	Indels 0;	Gaps 0

Qy	673	TCAGTCAGTATATCTGAGGTCCTGAGAACCTGTTGAGGTCGTCTGCTACTCAT	732
Dd	36258	TGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGA	36317
Qy	733	CCGGTGTGTGATTTGGTCAGACCTTGATTAATTAACAACGTGTTTGGATGCACTTTGTAC	792
Dd	36318	TGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTG	36377
Qy	733	AGTTCAGTAGCAGAGTGAATTTCCATCTCGGTCATTGTGTGGTGAAGTGAGGTGACGTGA	852
Dd	36378	TGTTTGTGTTGGGGGGGTTTGTTGTTTGGGTTGGGGGATGGGTGGGTGGGGG	36437
Qy	853	TGTGAGGTGAGTTGAGGTAGATTGGATGGGATGCAATGT	891
Dd	36438	TGTGTGTGTGTGTGGGTTGGGGGTTGGGAGGGGGGTGT	36476

Search completed: April 8, 2003, 08:50:51  
Job time : 6806 secs

**THIS PAGE BLANK (USPTO)**



FT /number= 2  
FT exon 1235..1477  
FT /\*tag= f  
FT /number= 3  
XX  
XX EP992582-A2.  
XX  
XX 12-APR-2000.  
XX  
XX 05-OCT-1999; 99EP-0307832.  
XX  
XX 07-OCT-1998; 98GB-0021821.  
XX  
XX (UYWA-) UNIV WALES BANGOR.  
XX  
XX Doenhoff M, Sayers J;  
XX WPI; 2000-259136/23.  
XX P-PSDB; AAY67587, AAY67588.  
XX  
XX New vaccine for treatment of Schistosoma infections contains a  
PT recombinant fusion protein comprising cercarial elastase sequence fused  
PT to bacterial, phage or viral protein -  
XX  
XX Claim 2, Fig 1; 26pp; English.  
XX  
XX The invention provides a vaccine comprising a recombinant fusion protein  
CC capable of eliciting immunity against Schistosoma parasites. The fusion  
CC protein comprises the 27 or 28 kDa cercarial elastase sequence of  
CC S. mansoni or an active fragment, homologue or variant, fused to a  
CC bacterial, phage or viral protein. The vaccine containing the fusion  
CC protein has been found to induce a significantly increased antibody  
CC response against schistosoma infections, compared to the use of S.  
CC mansoni cercarial elastase in its native form. The present sequence  
CC represents the S. mansoni cercarial elastase gene.  
XX  
XX Sequence 1477 BP; 348 A; 261 C; 402 G; 465 T; 1 other;  
SQ

Query Match 100.0%; Score 1476.6; DB 21; Length 1477;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 1477; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GTGTCAACCTGGTTGATAGTAGTGGAACCTGTGCAACACCGCACTGAAATCCGGTTC 60  
Db 1 GTGTCAACCTGGTTGATAGTAGTGGAACCTGTGCAACACCGCACTGAAATCCGGTTC 60  
QY 61 ATGCGATTCTTAAACGACAGAGAGAAACAATGTGTACAGSTTCACTAGTCTCAACGAGAGA 120  
Db 61 ATGCGATTCTTAAACGACAGAGAGAAACAATGTGTACAGSTTCACTAGTCTCAACGAGAGA 120  
QY 121 GTAAGTCAACAGCTGGTCAATGTGTGCTCAACCTGCGAAGTGGGGTAAAGATCGA 180  
Db 121 GTAAGTCAACAGCTGGTCAATGTGTGCTCAACCTGCGAAGTGGGGTAAAGATCGA 180  
QY 181 CTGAAACACATTTGCTCAATGTATTCATGATTTTCAACAGACAGTGTGTGGTGGC 240  
Db 181 CTGAAACACATTTGCTCAATGTATTCATGATTTTCAACAGACAGTGTGTGGTGGC 240  
QY 241 TGGCTTTGTGTTTGTATGACAGTGGGGTTTGCATTAATTTAATGTCTATATACCTGGTTC 300  
Db 241 TGGCTTTGTGTTTGTATGACAGTGGGGTTTGCATTAATTTAATGTCTATATACCTGGTTC 300  
QY 301 TTATTTCAAGTTTCATTTCTCACAACGAGAAATGGGCAACCAAGGCACTCCATCCAA 360  
Db 301 TTATTTCAAGTTTCATTTCTCACAACGAGAAATGGGCAACCAAGGCACTCCATCCAA 360  
QY 361 CCGTCTGAGATTAAAGTGGACCAAGATACACTCCCTCTTGATGTGCGACAGAGAG 420  
Db 361 CCGTCTGAGATTAAAGTGGACCAAGATACACTCCCTCTTGATGTGCGACAGAGAG 420  
QY 421 AGACCAATGCGACAGACACTCACTGATGATTTGATTTGCAATTTGATGCTGCTCAATG 480  
Db 421 AGACCAATGCGACAGACACTCACTGATGATTTGATTTGCAATTTGATGCTGCTCAATG 480

QY 481 GTCAACTTACAGAGTGGATATGAGATGATCACTCTGCCACAGCCATCGGATATCCCGCCA 540  
Db 481 GTCAACTTACAGAGTGGATATGAGATGATCACTCTGCCACAGCCATCGGATATCCCGCCA 540  
QY 541 CCGTGAACCTGGTTTTCATTTGTTATGGAAGGAGATGATTAACGACCGTATCCGTCA 600  
Db 541 CCGTGAACCTGGTTTTCATTTGTTATGGAAGGAGATGATTAACGACCGTATCCGTCA 600  
QY 601 CGTAAGAATGTGGAATATTGAAGAAAGTGAAGTTGTTGGAATTAACGACATGACTGAG 660  
Db 601 CGTAAGAATGTGGAATATTGAAGAAAGTGAAGTTGTTGGAATTAACGACATGACTGAG 660  
QY 661 TCGTCAGTCAAGTGTCACTGATTTCTGTGTGTCTGTGTATCTGTTTGTGTCTGTCT 720  
Db 661 TCGTCAGTCAAGTGTCACTGATTTCTGTGTGTCTGTGTATCTGTTTGTGTCTGTCT 720  
QY 721 TGTCTAAGTATCCGGTGTGTTGATATGTTGTCAGAGCTTGATATTAACAACGTGTGGA 780  
Db 721 TGTCTAAGTATCCGGTGTGTTGATATGTTGTCAGAGCTTGATATTAACAACGTGTGGA 780  
QY 781 TGACTTTGACAGTTCACTAGTACAGATGATTTCCATCTCGGTCAATGTGTGAGGT 840  
Db 781 TGACTTTGACAGTTCACTAGTACAGATGATTTCCATCTCGGTCAATGTGTGAGGT 840  
QY 841 GAGGTGACGTGATGTGAGGTGAGTTGAGGTGATTTGATGAGTGAATGTGATGTGATG 900  
Db 841 GAGGTGACGTGATGTGAGGTGAGTTGAGGTGATTTGATGAGTGAATGTGATGTGATG 900  
QY 901 GGATGATTTGACACCTCTTGAGAGAGAGAAAGTCAATGAAATTTCTATGCAACGATGGA 960  
Db 901 GGATGATTTGACACCTCTTGAGAGAGAGAAAGTCAATGAAATTTCTATGCAACGATGGA 960  
QY 961 GGTGTGTGTGTCATGAAAGTGAAGGTGATGATTTGTGAGTATGTTGAGAGTGTG 1020  
Db 961 GGTGTGTGTGTCATGAAAGTGAAGGTGATGATTTGTGAGTATGTTGAGAGTGTG 1020  
QY 1021 AGATGAGAGATGACTTGTATGCTGCAATATATGTCACATGTGATTTGATGTGACTAT 1080  
Db 1021 AGATGAGAGATGACTTGTATGCTGCAATATATGTCACATGTGATTTGATGTGACTAT 1080  
QY 1081 TGTGTGGGTAGTGTAAAGGTGATTTGTGCAAGTGTGATATTTTGCAGAAATCAGTTGG 1140  
Db 1081 TGTGTGGGTAGTGTAAAGGTGATTTGTGCAAGTGTGATATTTTGCAGAAATCAGTTGG 1140  
QY 1141 TGTGTTGTTTGTGCTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTT 1200  
Db 1141 TGTGTTGTTTGTGCTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTT 1200  
QY 1201 GTAAGTGTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTT 1260  
Db 1201 GTAAGTGTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTT 1260  
QY 1261 ACATGCGAACCAATGCGAATCTATATGTGTGAAGCAGGTGCAAGATTTGCGACAGTTACC 1320  
Db 1261 ACATGCGAACCAATGCGAATCTATATGTGTGAAGCAGGTGCAAGATTTGCGACAGTTACC 1320  
QY 1321 CCGTCCAGGTGACAGTGTGACCTCTCTCCATCCCTTCAAGGTCCAGTCACTGGTGT 1380  
Db 1321 CCGTCCAGGTGACAGTGTGACCTCTCTCCATCCCTTCAAGGTCCAGTCACTGGTGT 1380  
QY 1381 CGATACATGAGTGTGACACTGCACTCCAAACCTTCCGATATCATTTGTGAGTATGCCAGTGT 1440  
Db 1381 CGATACATGAGTGTGACACTGCACTCCAAACCTTCCGATATCATTTGTGAGTATGCCAGTGT 1440  
QY 1441 GGCTAGAAATGTTGAATTTGTAAGCTCCAAATATTGA 1477  
Db 1441 GGCTAGAAATGTTGAATTTGTAAGCTCCAAATATTGA 1477

RESULT 2  
ABI32679  
ID ABI32679 standard; DNA; 5474 BP.







[illegible]



XX The present invention provides a number of human immune system associated  
CC genes which are modified by the methylation of cytosines. The sequences  
CC can be used in the diagnosis and treatment of immune system disorders,  
CC including eye diseases such as retinopathy, neovascular glaucoma and  
CC macular degeneration, arteriosclerosis, anaemia, cancer, acute myeloid  
CC leukaemia, Alzheimer's disease, AIDS, epilepsy, neurofibromatosis,  
CC rheumatoid arthritis, psoriasis and inflammatory/ulcerative bowel  
CC diseases. The present sequence is a gene of the invention.

XX  
SQ Sequence 5680 BP; 1014 A; 231 C; 1956 G; 2479 T; 0 other;

Query Match 3.4%; Score 49.8; DB 24; Length 5680;  
Best Local Similarity 51.1%; Pred. No. 0.0015;  
Matches 117; Conservative 0; Mismatches 112; Indels 0; Gaps 0;

QY 990 AAGTGTTCAGATGATGTTGGAGAGTGGTGAATGAGAGTCAATTCGTCGAATAT 1049  
DB 4220 AAGATGTTACGATTTTATTTTGGAGGGGTGGAGTTAGCAAGTTGGCTTTGATTCG 4279  
QY 1050 AGTGCACATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1109  
DB 4280 AGTGTGTGTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTT 4339  
QY 1110 TGCCAGTTGATATTTTCGAAATTCACCTGCTGCTTTGTTTCCGTTGCTTTGTT 1169  
DB 4340 TGAAGGTTGTTGAGGTTTATTTATTTGCTTTTATTTGTTAGATTTGTTAGGTTTGT 4399  
QY 1170 CTTTCTCATCTGACTGATGATGATGATGATGATGATGATGATGATGATGATGATG 1218  
DB 4400 TTTTATTCGTTAGTTGAGTGAATGATGATGATGATGATGATGATGATGATGATGAT 4448

RESULT 9  
ABL33542  
ID ABL33542 standard; DNA; 6073 BP.

XX ABL33542;

XX 26-MAR-2002 (first entry)

XX Human immune system associated gene SEQ ID NO: 1515.

XX Human; immune system disease; cytosine methylation; antiasthmatic;  
KW antiarteriosclerotic; antianaemic; cytosolic; noctropic;  
KW neuroprotective; anti-HIV; anticonvulsant; ophthalmological;  
KW antineumatic; antiarthritic; antidiabetic; antipneumatic;  
KW antineumatic; antiarthritic; antidiabetic; antipneumatic;  
KW acute myeloid leukaemia; cancer; eye disease; arteriosclerosis; anaemia;  
KW neurofibromatosis; rheumatoid arthritis; psoriasis; bowel disease;  
KW gene; ds.

XX Homo sapiens.  
OS  
XX WO200200928-A2.  
XX  
XX 03-JAN-2002.  
XX  
XX 02-JUL-2001; 2001WO-EP07537.  
XX  
XX 30-JUN-2000; 2000DE-1032529.  
XX  
XX 01-SEP-2000; 2000DE-1043826.  
XX  
XX (EPIG-) EPIGENOMICS AG.

XX Olek A, Piepenbrock C, Berlin K;

XX WPI, 2002-130909/17.

XX Nucleic acid comprising fragment of chemically modified gene, useful  
PT for diagnosis and treatment of diseases associated with abnormal  
XX cytosine methylation -  
XX

PS Claim 1; SEQ ID NO 1515; 32pp + Sequence Listing; German.

XX The present invention provides a number of human immune system associated  
CC genes which are modified by the methylation of cytosines. The sequences  
CC can be used in the diagnosis and treatment of immune system disorders,  
CC including eye diseases such as retinopathy, neovascular glaucoma and  
CC macular degeneration, arteriosclerosis, anaemia, cancer, acute myeloid  
CC leukaemia, Alzheimer's disease, AIDS, epilepsy, neurofibromatosis,  
CC rheumatoid arthritis, psoriasis and inflammatory/ulcerative bowel  
CC diseases. The present sequence is a gene of the invention.

XX  
SQ Sequence 6073 BP; 1583 A; 51 C; 1289 G; 3150 T; 0 other;

Query Match 3.4%; Score 49.6; DB 24; Length 6073;  
Best Local Similarity 47.2%; Pred. No. 0.0017;  
Matches 151; Conservative 0; Mismatches 169; Indels 0; Gaps 0;

QY 906 ATTGAGACCACTGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 965  
DB 3300 ATTGCGAATTTAGATATTAATTAAGAGATATGATATTTAGTGAATAATTTTC 3359  
QY 966 TTGTGTACATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1025  
DB 3360 GTGAAGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1025  
QY 1026 GAGAGTACCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1085  
DB 3420 GTGATTTAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3479  
QY 1086 GGGATGATGAAAGGGTGAATTTGTCAGATGATGATGATGATGATGATGATGATGAT 1145  
DB 3480 GGGAGTGTGTTTATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 3539  
QY 1146 TGTGTTGTCGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTG 1205  
DB 3540 TAAATTTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 3599  
QY 1206 GTTGTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTG 1225  
DB 3600 TTTTATTTAGTTTGTGTT 3619

RESULT 10  
ABL32912  
ID ABL32912 standard; DNA; 6621 BP.

XX ABL32912;

XX 26-MAR-2002 (first entry)

XX Human immune system associated gene SEQ ID NO: 865.

XX Human; immune system disease; cytosine methylation; antiasthmatic;  
KW antiarteriosclerotic; antianaemic; cytosolic; noctropic;  
KW neuroprotective; anti-HIV; anticonvulsant; ophthalmological;  
KW antineumatic; antiarthritic; antidiabetic; antipneumatic;  
KW antineumatic; antiarthritic; antidiabetic; antipneumatic;  
KW acute myeloid leukaemia; cancer; eye disease; arteriosclerosis; anaemia;  
KW neurofibromatosis; rheumatoid arthritis; psoriasis; bowel disease;  
KW gene; ds.

XX Homo sapiens.  
OS  
XX WO200200928-A2.  
XX  
XX 03-JAN-2002.  
XX  
XX 02-JUL-2001; 2001WO-EP07537.  
XX  
XX 30-JUN-2000; 2000DE-1032529.  
XX  
XX 01-SEP-2000; 2000DE-1043826.  
XX  
XX (EPIG-) EPIGENOMICS AG.

XX Olek A, Piepenbrock C, Berlin K;

XX WPI, 2002-130909/17.

XX Nucleic acid comprising fragment of chemically modified gene, useful  
PT for diagnosis and treatment of diseases associated with abnormal  
XX cytosine methylation -  
XX



KM ATR-X; Bloom's syndrome; tumour; cancer; methylation; gene; ds.  
XX Homo sapiens.  
XX WO200177377-A2.  
XX 18-OCT-2001.  
XX 06-APR-2001; 2001WO-EP03971.  
XX 06-APR-2000; 2000DE-1019058.  
XX 07-APR-2000; 2000DE-1019173.  
XX 30-JUN-2000; 2000DE-1032529.  
XX 01-SEP-2000; 2000DE-1043826.  
XX (EPIC-) EPIGENOMICS AG.  
XX Olek A, Piepenbrock C, Berlin K;  
XX WPI; 2002-017471/02.  
XX New nucleic acid sequences from chemically modified genes associated  
PT with DNA replication, useful for analysing cytosine methylations for  
PT diagnosis and therapy of diseases e.g. Ataxia telangiectasia -  
XX Claim 1; SEQ ID NO 72; 23pp + Sequence Listing; English.  
XX  
XX The invention relates to nucleic acid sequences comprising at least 18  
CC bases of a chemically pretreated gene associated with gene regulation,  
CC selected from 94 genes (ABL49301-ABL49394) and/or complementary sequences  
CC associated with DNA replication, CENPB, DNAL2, ATR, CHD1L, ERCC3, SNRPAL,  
CC RAD50 and LIG2. The chemical pretreatment converts cytosine bases  
CC unmethylated at the 5-position to uracil or another base with  
CC hybridisation behaviour dissimilar to cytosine, to enable analysis of  
CC cytosine methylations. The DNA sequences and method are useful in the  
CC diagnosis of diseases (or predisposition to diseases) associated with DNA  
CC replication and in therapy of such diseases, by enabling analysis of the  
CC cytosine methylation patterns of such genes. They are especially useful  
CC in diagnosis and therapy of e.g. Ataxia telangiectasia, ATR-X, Bloom's  
CC syndrome, solid tumours and cancer.  
CC Note: The sequence data for this patent did not form part of the printed  
CC specification and was supposed to be available directly from WIPO at  
CC ftp.wipo.int/pub/published\_pat\_sequences. However, the sequence data did  
CC not correspond to that referred to in the specification. The present data  
CC is taken from EPO data for the patent.  
XX  
XX Sequence 8607 BP; 2442 A; 146 C; 1904 G; 4115 T; 0 other;  
SQ  
Query Match 3.3%; Score 49.4; DB 24; Length 8607;  
Best Local Similarity 56.4%; Pred. No. 0.0023;  
Matches 92; Conservative 0; Mismatches 71; Indels 0; Gaps 0;  
QY 1063 TTGATGCGACTATTGTTGTGGGTAGCTGTAAGGGTGAATATTGTCAGTGTGAT 1122  
DB 6933 TTGATGCGTTTATGTTATGTTTATTTATTTATGTTATTTTATTTT 6992  
QY 1123 TTGCAAAATCTGCTGCTGTTTGTTCGCGTGGTGTGCTTCTTCATCTGT 1182  
DB 6993 TTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTT 7052  
QY 1183 ACTGACGTTGTTGACGTAAGTGTGTTGTTGTTGTTT 1225  
DB 7053 TTTGTTTTTTTTTTTTTTTTTTTTTTTTTTT 7095  
RESULT 13  
ABL33119  
ID ABL33119 standard; DNA; 8758 BP.  
XX  
XX ABL33119;  
XX  
XX 26-MAR-2002 (first entry)  
XX

DE Human immune system associated gene SEQ ID NO: 1092.  
XX  
XX Human; immune system disease; cytosine methylation; antiasthmatic;  
XX antiarteriosclerotic; antianemic; cytostatic; nootropic;  
XX neuroprotective; anti-HIV; anticonvulsant; ophthalmologic;  
XX antirheumatic; antiarthritic; antidiabetic; antiporiatic;  
XX antinflammatory; cancer; eye disease; arteriosclerosis; anaemia;  
XX acute myeloid leukaemia; Alzheimer's disease; AIDS; epilepsy;  
XX neurofibromatosis; rheumatoid arthritis; psoriasis; bowel disease;  
XX gene; ds.  
XX  
XX Homo sapiens.  
XX WO200200928-A2.  
XX 03-JAN-2002.  
XX 02-JUN-2001; 2001WO-EP07537.  
XX 30-JUN-2000; 2000DE-1032529.  
XX 01-SEP-2000; 2000DE-1043826.  
XX (EPIC-) EPIGENOMICS AG.  
XX Olek A, Piepenbrock C, Berlin K;  
XX WPI; 2002-130909/17.  
XX  
XX Nucleic acid comprising fragment of chemically modified gene, useful  
PT for diagnosis and treatment of diseases associated with abnormal  
PT cytosine methylation -  
XX  
XX Claim 1; SEQ ID NO 1092; 32pp + Sequence Listing; German.  
XX  
XX The present invention provides a number of human immune system associated  
CC genes which are modified by the methylation of cytosines. The sequences  
CC can be used in the diagnosis and treatment of immune system disorders,  
CC including eye diseases such as retinopathy, neovascular glaucoma and  
CC macular degeneration, arteriosclerosis, anaemia, cancer, acute myeloid  
CC leukaemia, Alzheimer's disease, AIDS, epilepsy, neurofibromatosis,  
CC rheumatoid arthritis, psoriasis and inflammatory/bowel  
CC diseases. The present sequence is a gene of the invention.  
XX  
XX Sequence 8758 BP; 2209 A; 217 C; 2178 G; 4154 T; 0 other;  
SQ  
Query Match 3.3%; Score 49.2; DB 24; Length 8758;  
Best Local Similarity 49.2%; Pred. No. 0.0027;  
Matches 129; Conservative 0; Mismatches 133; Indels 0; Gaps 0;  
QY 964 TGTGTGACATGAAAGTGGGGGTCATGCTGTTGAGTATGTTGAGAGTGAGTA 1023  
DB 6183 TGTGTTAGCTTTAGTGTAAGGGGTATATTTTGTGGGTAGCTTTTGTAGTATGTAAGG 6242  
QY 1024 TGAAGATGACTTGATCGTCGAATATAGTGCACATGATGTAATGAGCTATTGTT 1083  
DB 6243 TTATTAAGATTTGTTATTTTATTTTATTTAGTGAATTTTATTAATGTAATTTTGT 6302  
QY 1084 GTGGGTAGTGAAGGGGTGAATTTGTCAGCTGATATTTTCAAAATTCACCTGTGT 1143  
DB 6303 GATATTTTATGATGATGTTAAGATTTTGAAGATTTATTTATTTTATTTGAGATT 6362  
QY 1144 TTTGTTTGTCCGTTGCTGTTTGTCTTCTCATCTGTAAGCTGTTGTAAGCTGA 1203  
DB 6363 TTTAATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTGA 6422  
QY 1204 CTGTTGTGCGTTGTTGTTT 1225  
DB 6423 ATGTTTTTTTATGAATTTT 6444  
RESULT 14  
ABL70185  
ID ABL70185 standard; DNA; 4661 BP.  
XX

XX ABL70185;  
AC  
XX 01-JUN-2002 (first entry)  
DT  
XX Chemically treated cell signalling DNA sequence#38.  
DE  
XX Cell signalling; cytosine methylation; cell signalling disease;  
KW cancer; tumour; cytostatic; ds.  
XX  
XX Unidentified.  
OS  
XX WO200202807-A2.  
PN  
XX 10-JAN-2002.  
PD  
XX 29-JUN-2001; 2001WO-EP07471.  
PE  
XX 30-JUN-2000; 2000DE-1032529.  
PR  
XX 01-SEP-2000; 2000DE-1043826.  
PR  
XX (EPIC-) EPIGENOMICS AG.  
PA  
XX Olek A, Piepenbrock C, Berlin K;  
PI  
XX WPI; 2002-154758/20.  
DR  
XX  
XX Nucleic acid, useful for diagnosis and therapy of diseases associated  
PT with cell signalling e.g. cancer, comprises chemically modified genomic  
PT sequences of genes associated with cell signalling -  
XX  
XX Claim 1; SEQ ID NO 75; 24p+sequence listing; English.  
PS  
XX  
XX The invention relates to a nucleic acid comprising a sequence of at least  
CC 18 bases of a segment of chemically pretreated DNA of genes associated  
CC with cell signalling. The activity of the modified sequences of the  
CC invention may be described as cytostatic. The object of the invention is  
CC to provide the chemically modified DNA of genes associated with cell  
CC signalling, as well as oligonucleotides and/or PNA-oligomers for  
CC detecting cytosine methylations, as well as a method which is  
CC particularly suitable for the diagnosis and/or therapy of genetic and  
CC epigenetic parameters of genes associated with cell signalling. The  
CC chemically modified DNA provided by the invention is useful for diagnosis  
CC and therapy of diseases such as solid tumours and cancer. The sequences  
CC given in records ABL70111-ABL70626 represent chemically pre-treated  
CC genomic DNA's of genes associated with cell signalling  
CC Note: The sequence data for this patent is not represented in the printed  
CC specification, but is based on sequence information supplied by the  
CC European Patent Office.  
CC  
SQ Sequence 4661 BP; 821 A; 98 C; 1300 G; 2442 T; 0 other;  
Query Match 3.3%; Score 49; DB 24; Length 4661;  
Best Local Similarity 50.6%; Pred. No. 0.0022;  
Matches 118; Conservative 0; Mismatches 115; Indels 0; Gaps 0;

RESULT 15  
AAS61136  
ID AAS61136 standard; DNA; 4661 BP.  
XX  
XX AAS61136;  
AC  
XX 29-JAN-2002 (first entry)  
DT  
XX Human gene regulation-associated gene oligonucleotide #91.  
DE  
XX  
XX Human; Gene regulation-associated gene; severe combined immunodeficiency;  
KW cardiac damage; inflammatory response; Haemophilia; Werner syndrome;  
KW asthma; HDR syndrome; congenital heart defect; Saethre-Chotzen syndrome;  
KW renal disease; Preclampsia; cardiac allograft vascular disease;  
KW colorectal cancer; thyroid cancer; oesophageal cancer; ds; tumour;  
KW immunostimulant; cardiant; antiinflammatory; coagulant; antiasthmatic;  
KW nephrotropic; gynecological; anti-tumour; immunosuppressive; cytostatic.  
XX  
XX Homo sapiens.  
OS  
XX WO20017375-A2.  
PN  
XX 18-OCT-2001.  
PD  
XX 06-APR-2001; 2001WO-EP03968.  
PE  
XX 06-APR-2000; 2000DE-1019058.  
PR  
XX 07-APR-2000; 2000DE-1031973.  
PR  
XX 30-JUN-2000; 2000DE-1032529.  
PR  
XX 01-SEP-2000; 2000DE-1043826.  
PR  
XX (EPIC-) EPIGENOMICS AG.  
PA  
XX Olek A, Piepenbrock C, Berlin K;  
PI  
XX WPI; 2002-017470/02.  
DR  
XX  
XX New nucleic acid sequences from chemically modified genes associated  
PT with gene regulation, useful for analysing cytosine methylations for  
PT diagnosis and therapy of diseases e.g. severe combined immunodeficiency  
PT disease -  
XX  
XX Claim 1; SEQ ID NO 93; 26p; English.  
PS  
XX  
XX The invention relates to 224 nucleic acid sequences comprising at least  
CC 18 bases of a chemically pretreated gene associated with gene regulation  
CC selected from 43 known genes (or complementary sequences). The  
CC chemical pretreatment converts cytosine bases unmethylated at the  
CC 5-position to uracil or another base with hybridisation behaviour  
CC dissimilar to cytosine, to enable analysis of cytosine methylations.  
CC The DNA sequences, oligomers (or sets/arrays) and method are  
CC useful in the diagnosis of diseases (or predisposition to diseases)  
CC associated with gene regulation and in therapy of such diseases, by  
CC enabling analysis of the cytosine methylation patterns of such genes,  
CC kits are provided. They are especially useful in diagnosis  
CC and therapy of e.g. severe combined immunodeficiency disease, cardiac  
CC disorders, haemophilia, solid tumours and cancer, Werner syndrome,  
CC asthma, HDR syndrome, Saethre-Chotzen syndrome, renal disease,  
CC preclampsia, graft versus-host disease. The present sequence is a  
CC sequence included in the sequence data for this specification and is  
CC associated with the human gene regulation-associated genes.  
CC Note: The sequence data for this patent did not form part  
CC of the printed specification, but was obtained in electronic  
CC format directly from WIPO at  
CC ftp.wipo.int/pub/published\_pct\_sequences  
XX  
SQ Sequence 4661 BP; 821 A; 98 C; 1300 G; 2442 T; 0 other;  
Query Match 3.3%; Score 49; DB 24; Length 4661;  
Best Local Similarity 50.6%; Pred. No. 0.0022;  
Matches 118; Conservative 0; Mismatches 115; Indels 0; Gaps 0;

Db 3644 TTTTGGTGGTGGGCGGGGAGGGGTTAGGTTTGTCTTTTAAAGG 3703  
 Oy 1053 GCACATGATGATGATGACATATGTTGCGGTAGTGAAGCGGATATGTC 1112  
 Db 3704 TTTAATGTTTGGTTTGTCTTTTTCGTTTATGTTGTTGTTGTTT 3763  
 Oy 1113 CAGTTGATATTTTGCAGAACTGTTGTTTGTTCCTGTTGTTTGTCTT 1172  
 Db 3764 TGGTTTATTTTTCGGAATTTTTCGGAATTTTTCGGAATTTTATGTTT 3823  
 Oy 1173 TCTCATCTGACTGACGTTGTTGACGTTACTGTTGTTGTTGTTT 1225  
 Db 3824 TTTTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 3876

# RESULT 16 ABR31226

ID ABR31226 standard; DNA; 4661 BP.

AC ABR31226;

DT 23-APR-2002 (first entry)

DE Signal transduction associated gene modified DNA #35.

KW Human; signal transduction associated gene; cytosine methylation strate;

KW CpG island; signal transduction associated disease; solid tumour; cancer;

KW antitumour; cytostatic; mutant; ds.

XX Homo sapiens.

OS Synthetic.

PN WO200200926-A2.

PD 03-JAN-2002.

PF 29-JUN-2001; 2001WO-EP07472.

PR 30-JUN-2000; 2000DE-1032529.

PR 01-SEP-2000; 2000DE-1043826.

XX (EPIG-) EPIGENOMICS AG.

XX Olek A, Piepenbrock C, Berlin K;

PI WPI; 2002-147896/19.

DR

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

CC European Patent Office.  
 XX SQ Sequence 4661 BP; 821 A; 98 C; 1300 G; 2442 T; 0 other;  
 Query Match 3.3%; Score 49; DB 24; Length 4661;  
 Best Local Similarity 50.6%; Pred. No. 0.0022;  
 Matches 118; Conservative 0; Mismatches 115; Indels 0; Gaps 0;

Oy 993 TGTGTGAGTATGTTTGGAGAGTGTGAGATGAGAGTGAATGATGATATAGT 1052  
 Db 3644 TTTTGGTGGTGGGCGGGGAGGGGTTAGGTTTGTCTTTTAAAGG 3703  
 Oy 1053 GCACATGATGATGATGACATATGTTGCGGTAGTGAAGCGGATATGTC 1112  
 Db 3704 TTTAATGTTTGGTTTGTCTTTTTCGTTTATGTTGTTGTTTGT 3763  
 Oy 1113 CAGTTGATATTTTGCAGAACTGTTGTTTGTTCCTGTTGTTTGTCTT 1172  
 Db 3764 TGGTTTATTTTTCGGAATTTTTCGGAATTTTTCGGAATTTTATGTTT 3823  
 Oy 1173 TCTCATCTGACTGACGTTGTTGACGTTACTGTTGTTGTTGTTT 1225  
 Db 3824 TTTTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 3876

# RESULT 17 ABL34427

ID ABL34427 standard; DNA; 9209 BP.

AC ABL34427;

DT 26-MAR-2002 (first entry)

DE Human immune system associated gene SEQ ID NO: 2400.

KW Human; immune system disease; cytosine methylation; antiasthmatic;

KW antiarteriosclerotic; antianemic; cytostatic; noctropic;

KW neuroprotective; anti-HIV; anticonvulsant; ophthalmological;

KW antirheumatic; antiarthritic; antidiabetic; antipsoriatic;

KW antiinflammatory; cancer; eye disease; arteriosclerosis; anaemia;

KW acute myeloid leukemia; Alzheimer's disease; AIDS; epilepsy;

KW neurofibromatosis; rheumatoid arthritis; psoriasis; bowel disease;

XX gene; ds.

XX Homo sapiens.

OS WO200200928-A2.

PN 03-JAN-2002.

PD 02-JUL-2001; 2001WO-EP07537.

PF 30-JUN-2000; 2000DE-1032529.

PR 01-SEP-2000; 2000DE-1043826.

XX (EPIG-) EPIGENOMICS AG.

XX Olek A, Piepenbrock C, Berlin K;

PI WPI; 2002-130909/17.

DR

XX

XX

XX

XX

The present invention provides a number of human immune system associated genes which are modified by the methylation of cytosines. The sequences can be used in the diagnosis and treatment of immune system disorders, including eye diseases such as retinopathy, neovascular glaucoma and macular degeneration, arteriosclerosis, anaemia, cancer, acute myeloid leukaemia, Alzheimer's disease, AIDS, epilepsy, neurofibromatosis,

CC rheumatoid arthritis, psoriasis and inflammatory/ulcerative bowel  
CC diseases. The present sequence is a gene of the invention.  
XX  
SQ Sequence 9209 BP; 2206 A; 126 C; 2464 G; 4413 T; 0 other;

Query Match 3.3%; Score 49; DB 24; Length 9209;  
Best Local Similarity 50.2%; Pred. No. 0.0031;  
Matches 121; Conservative 0; Mismatches 120; Indels 0; Gaps 0;

Qy 985 GGTCAATGTTGATGATGTTGGAGAGTGGAGATGAGAGTGAATGATGATGCTG 1044  
Db 22 GGTATATATCGAGGAGATATTTTAAAGAGAAATAGTTGGTTGTTGGTGGC 81  
Qy 1045 AATATAGTCACATGATGATGATGATGATGATGATGATGATGATGATGATG 1104  
Db 82 TTTTATAGTTTGAAGGATGTTGTTTATTTTATTTATTTGATATTTATTTTACG 141  
Qy 1105 TATGTGCGCAGTTGATTTTGAATTCACCTGTTGTTGTTGTTGTTGTTGTT 1164  
Db 142 TATTTAATAGTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTT 201  
Qy 1165 TTGTTCTTTCATCTGATCTGATGATGATGATGATGATGATGATGATGATGAT 1224  
Db 202 TTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTT 261  
Qy 1225 T 1225  
Db 262 T 262

RESULT 18  
AAS46808  
ID AAS46808 standard; DNA; 9219 BP.

AC AAS46808;

DT 18-DEC-2001 (first entry)

DE Tumour suppressor gene derived chemically modified sequence #534.

XX Human; tumour suppressor gene; oncogene; antitumour; cytostatic;  
KM cancer; tumour; CpG dinucleotide; single-nucleotide polymorphism; SNP;  
KW cytosine methylation; ds.

OS Homo sapiens.

PN WO200168912-A2.

PD 20-SEP-2001.

PF 15-MAR-2001; 2001WO-EP02955.

PR 15-MAR-2000; 2000DE-1013847.

PR 06-APR-2000; 2000DE-1013058.

PR 07-APR-2000; 2000DE-1019173.

PR 30-JUN-2000; 2000DE-1032529.

PR 01-SEP-2000; 2000DE-1043826.

XX (EPIC-) EPIGENOMICS AG.

PI Olek A, Plepenbrock C, Berlin K;

DR WPI, 2001-602752/68.

XX Fragments of chemically modified genes associated with tumour suppressor  
PT genes and oncogenes, useful in designing primers and probes for  
PT analysing diseases associated with cytosine methylation state e.g.  
cancer -  
PS Claim 1, SEQ ID No 534; 27pp; English.

XX The invention relates to a nucleic acid comprising a sequence of 18  
CC bases, of a segment of chemically pretreated DNA (CP DNA) e.g. with

CC bisulphite, of genes associated with tumour suppression and  
CC sequences having a sequence taken from 536 (actually 533 since  
CC numbers 408, 458 and 500 are missing from the sequence listing) sequences  
CC (Ss) and sequences complementary to (Ss). The nucleic acid may be a  
CC peptide nucleic acid-oligomer (PNA) of at least 9 nucleotides and may  
CC form part of a set of probes for detecting the cytosine methylation state  
CC and/or single nucleotide polymorphisms and also to be used in an  
CC array for analysing diseases associated with CpG dinucleotides e.g.  
CC cancers and tumours. The probes can also be used in a method for  
CC ascertaining genetic and/or epigenetic parameters for the diagnosis  
CC and/or therapy of existing diseases or the predisposition to specific  
CC diseases, by analysing cytosine methylations. The parameters may be  
CC compared to another set of genetic and/or epigenetic parameters, the  
CC differences serving as basis for diagnosis and/or prognosis events which  
CC are disadvantageous to patients. The present sequence is one of the  
CC 533 genomic sequences derived from tumour suppressor genes and  
CC oncogenes. Sequences with even numbered Seq ID numbers are the  
CC complementary sequence of the corresponding odd numbered sequence (e.g.  
CC ID 2 and ID1, ID 536 and ID 535, except for those whose partner sequence  
CC is missing).  
CC Note: The sequence data for this patent did not form part  
CC of the printed specification, but was obtained in electronic  
CC format directly from WIPO at  
CC ftp.wipo.int/pub/published\_pct\_sequences.

XX Sequence 9219 BP; 2206 A; 126 C; 2470 G; 4417 T; 0 other;

Query Match 3.3%; Score 49; DB 22; Length 9219;  
Best Local Similarity 50.2%; Pred. No. 0.0031;  
Matches 121; Conservative 0; Mismatches 120; Indels 0; Gaps 0;

Qy 985 GGTCAATGTTGATGATGTTGGAGAGTGGAGATGAGAGTGAATGATGATGCTG 1044  
Db 22 GGTATATATCGAGGAGATATTTTAAAGAGAAATAGTTGGTTGTTGGTGGC 81  
Qy 1045 AATATAGTCACATGATGATGATGATGATGATGATGATGATGATGATGATG 1104  
Db 82 TTTTATAGTTTGAAGGATGTTGTTTATTTTATTTTATTTTATTTTATTTTATTT 141  
Qy 1105 TATGTGCGCAGTTGATTTTGAATTCACCTGTTGTTGTTGTTGTTGTTGTT 1164  
Db 142 TATTTAATAGTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTT 201  
Qy 1165 TTGTTCTTTCATCTGATCTGATGATGATGATGATGATGATGATGATGATGAT 1224  
Db 202 TTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTT 261  
Qy 1225 T 1225  
Db 262 T 262

RESULT 19  
AAS45399  
ID AAS45399 standard; DNA; 15832 BP.

AC AAS45399;

DT 18-DEC-2001 (first entry)

DE Chemically pretreated complementary DNA associated with cell cycle #52.

XX Cell cycle; human; CpG dinucleotide; cytosine methylation; HIV; aging;  
KM human immunodeficiency virus; neurodegenerative disorder; solid tumour;  
KM graft-versus-host disease; glomerular disease; Levy body disease; cancer;  
KM arthritis; arteriosclerosis; anti-HIV; neuroprotective; antiarthritic;  
KM immunosuppressive; antitumour; cytostatic; antiarteriosclerotic; ds;  
KM PCR primer.

OS Homo sapiens.

PN WO200168911-A2.

PD 20-SEP-2001.  
 XX  
 PF 15-MAR-2001; 2001WO-EP02945.  
 XX  
 PR 15-MAR-2000; 2000DE-1013847.  
 PR 06-APR-2000; 2000DE-1019058.  
 PR 07-APR-2000; 2000DE-1019173.  
 PR 30-JUN-2000; 2000DE-1032529.  
 PR 01-SEP-2000; 2000DE-1043826.  
 XX  
 PA (EPIC-) EPIGENOMICS AG.  
 XX  
 PI Olek A, Piepenbrock C, Berlin K;  
 XX  
 DR WPI; 2001-602751/68.  
 XX  
 PT Designing primers and probes for analysing diseases associated with  
 PT cytosine methylation state e.g. arthritis, cancer, aging,  
 PT arteriosclerosis comprising fragments of chemically modified genes  
 PT associated with cell cycle -  
 XX  
 PS Claim 1; SEQ ID NO 104; 28pp; English.  
 XX  
 CC Sequences AAS45296-AAS45520 represent chemically pretreated genomic DNA  
 CC molecules associated with the cell cycle and specific PCR primers of the  
 CC invention. The sequences are useful for detecting the methylation state  
 CC of all CpG dinucleotides in a sequence and therefore for analysing  
 CC associated diseases. By analysing cytosine methylations in the pretreated  
 CC DNA, genetic and/or epigenetic parameters for the diagnosis and therapy  
 CC of existing diseases or the predisposition to specific diseases can be  
 CC ascertained. The parameters may be compared to another set of genetic  
 CC and/or epigenetic parameters, the differences serving as basis for  
 CC diagnosis and/or prognosis events which are disadvantageous to patients.  
 CC The sequences of the invention are useful for the diagnosis and therapy  
 CC of HIV infection, neurodegenerative disorders, graft-versus-host disease,  
 CC aging, glomerular disease, Lewy body disease, arthritis,  
 CC arteriosclerosis, solid tumours and cancers.  
 CC  
 XX  
 SQ Sequence 15832 BP; 4164 A; 297 C; 3742 G; 7610 T; 19 other;  
 Query Match 3.3%; Score 49; DB 22; Length 15832;  
 Best Local Similarity 48.7%; Pred. No. 0.004;  
 Matches 133; Conservative 0; Mismatches 140; Indels 0; Gaps 0;  
 Oy 953 CGATGAGTGTGTTGTGATCAATGAGGGGCAATGTTGATGATGTTGA 1012  
 Db 4409 CATTAGAGATTGTTGGTTAAATGATTAATTCGATTTAGTGAATTTATTATAC 4468  
 Oy 1013 GAGTGTGATGAGAGATGATCTGATCGAATATAGTGAATGATGATG 1072  
 Db 4469 GAGTTTAAAGAAATTTTAAATTTTAAATTTTAAATTTTGAATTTTGA 4528  
 Oy 1073 ACTATTGTTGTGGTGATGTAAGGGTGATATTGTCAGTTGATTTTGA 1132  
 Db 4529 TAAATTTTATTTTATTTTGAACGTTTATTTTATTTTATTTTATTTT 4588  
 Oy 1133 CACTGTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTT 1192  
 Db 4589 TTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTT 4648  
 Oy 1193 GTTGACTGACTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTT 1225  
 Db 4649 TTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTT 4681  
 RESULT 20  
 ABL33343  
 ID ABL33343 standard; DNA; 15832 BP.  
 XX  
 AC ABL33343;  
 XX  
 DT 26-MAR-2002 (first entry)  
 XX

DE Human immune system associated gene SEQ ID NO: 1316.  
 XX  
 KW Human; immune system disease; cytosine methylation; antiasthmatic;  
 KW antiarteriosclerotic; anti-nausea; cytosolic; nocotropic;  
 KW neuroprotective; anti-HIV; anticonvulsant; ophthalmological;  
 KW antidiabetic; antidiabetic; cancer; eye disease; arteriosclerosis; anaemia;  
 KW antiinflammatory; cancer; eye disease; arteriosclerosis; anaemia;  
 KW acute myeloid leukaemia; Alzheimer's disease; AIDS; epilepsy;  
 KW neurofibromatosis; rheumatoid arthritis; psoriasis; bowel disease;  
 KW gene; ds.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO200200928-A2.  
 XX  
 PD 03-JAN-2002.  
 XX  
 PF 02-JUN-2001; 2001WO-EP07537.  
 XX  
 PR 30-JUN-2000; 2000DE-1032529.  
 PR 01-SEP-2000; 2000DE-1043826.  
 XX  
 PA (EPIC-) EPIGENOMICS AG.  
 XX  
 PI Olek A, Piepenbrock C, Berlin K;  
 XX  
 DR WPI; 2002-130909/17.  
 XX  
 PT Nucleic acid comprising fragment of chemically modified gene, useful  
 PT for diagnosis and treatment of diseases associated with abnormal  
 PT cytosine methylation -  
 XX  
 PS Claim 1; SEQ ID NO 1316; 32pp + Sequence Listing; German.  
 XX  
 CC The present invention provides a number of human immune system associated  
 CC genes which are modified by the methylation of cytosines. The sequences  
 CC can be used in the diagnosis and treatment of immune system disorders,  
 CC including eye diseases such as retinopathy, neovascular glaucoma and  
 CC macular degeneration, arteriosclerosis, anaemia, cancer, acute myeloid  
 CC leukaemia, Alzheimer's disease, AIDS, epilepsy, neurofibromatosis,  
 CC rheumatoid arthritis, psoriasis and inflammatory/ulcerative bowel  
 CC diseases. The present sequence is a gene of the invention.  
 CC  
 XX  
 SQ Sequence 15832 BP; 4164 A; 297 C; 3742 G; 7610 T; 19 other;  
 Query Match 3.3%; Score 49; DB 24; Length 15832;  
 Best Local Similarity 48.7%; Pred. No. 0.004;  
 Matches 133; Conservative 0; Mismatches 140; Indels 0; Gaps 0;  
 Oy 953 CGATGAGTGTGTTGTGATCAATGAGGGGCAATGTTGATGATGTTGA 1012  
 Db 4409 CATTAGAGATTGTTGGTTAAATGATTAATTCGATTTAGTGAATTTATTATAC 4468  
 Oy 1013 GAGTGTGATGAGAGATGATCTGATCGAATATAGTGAATGATGATG 1072  
 Db 4469 GAGTTTAAAGAAATTTTAAATTTTAAATTTTAAATTTTGAATTTTGA 4528  
 Oy 1073 ACTATTGTTGTGGTGATGTAAGGGTGATATTGTCAGTTGATTTTGA 1132  
 Db 4529 TAAATTTTATTTTATTTTGAACGTTTATTTTATTTTATTTTATTTT 4588  
 Oy 1133 CACTGTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTT 1192  
 Db 4589 TTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTT 4648  
 Oy 1193 GTTGACTGACTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTT 1225  
 Db 4649 TTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTT 4681  
 RESULT 21  
 ABR28244  
 ID ABR28244 standard; DNA; 15832 BP.  
 XX

XX ABR28244;  
AC  
XX 23-APR-2002 (first entry)  
XX  
XX DNA transcription associated complementary genomic DNA #59.  
XX  
XX DNA transcription associated gene; peptide nucleic acid; PNA-oligomer;  
XX PNA; cytosine methylation state; SNP; retroviral infection; gene; ds;  
XX single nucleotide polymorphism; adenosine deaminase deficiency; cancer;  
XX viral infection; Sezary syndrome; haematological disorder; tuberculosis;  
XX immunological disorder; Werner syndrome; developmental disorder;  
XX psoriasis; Kieger's syndrome; neurological disorder; erythropoiesis;  
XX neurodegenerative disorder; Maardenburg syndrome; Niemann-Pick disease;  
XX myelodysplastic syndrome; myocardial infarction; hypertension; arthritis;  
XX angiogenesis; congenital heart disease; HDR syndrome; gene therapy;  
XX polyglutamine disorder; solid tumour.  
XX  
XX Unidentified.  
XX  
XX WO200192565-A2.  
XX  
XX 06-DEC-2001.  
XX  
XX 06-APR-2001; 2001WO-EP03973.  
XX  
XX 06-APR-2000; 2000DE-1019058.  
XX 07-APR-2000; 2000DE-1019173.  
XX 30-JUN-2000; 2000DE-1032529.  
XX 01-SEP-2000; 2000DE-1043826.  
XX  
XX (EPIG-) EPIGENOMICS AG.  
XX  
XX Olek A, Piepenbrock C, Berlin K;  
XX  
XX WPI; 2002-090046/12.  
XX  
XX New nucleic acids or oligomers, useful for diagnosing or treating  
XX diseases associated with DNA transcription, e.g. immunological  
XX disorders, Werner syndrome, psoriasis, myocardial infarction, solid  
XX tumours or cancer -  
XX  
XX Claim 1; SEQ ID NO 118; 32pp; English.  
XX  
XX The invention relates to a nucleic acid, which comprises a segment of the  
XX chemically pretreated DNA of genes associated with DNA transcription from  
XX one of 346 sequences, and an oligomer, in particular an oligonucleotide  
XX or peptide nucleic acid (PNA)-oligomer that hybridises to or is identical  
XX to the chemically pretreated DNA of genes associated with DNA  
XX transcription. The set of oligomer probes are useful for detecting the  
XX cytosine methylation state and/or single nucleotide polymorphisms (SNPs)  
XX in a chemically pretreated genomic DNA. The nucleic acids are useful for  
XX diagnosing or treating diseases associated with DNA transcription  
XX (particularly with the methylation status), e.g. adenosine deaminase  
XX deficiency, viral infection, retroviral infection, Sezary syndrome,  
XX haematological disorders, immunological disorders, Werner syndrome,  
XX tuberculosis, developmental disorders, psoriasis, Kieger's syndrome,  
XX neurological disorders, neurodegenerative disorders, Maardenburg  
XX syndrome, Niemann-Pick disease, myelodysplastic syndrome, myocardial  
XX infarction, hypertension, angiogenesis, erythropoiesis, congenital heart  
XX disease, HDR syndrome, arthritis, polyglutamine disorders, solid tumours  
XX or cancer. Sequences ABR28127-ABR28472 represent DNA transcription  
XX associated genomic DNA molecules of the invention.  
XX Note: The sequence data for this patent did not form part of the printed  
XX specification but was obtained in electronic format directly from the  
XX European Patent Office.  
XX  
XX Sequence 15832 BP; 4164 A; 297 C; 3742 G; 7610 T; 19 other;  
XX  
XX Query Match 3.3%; Score 49; DB 24; Length 15832;  
XX Best Local Similarity 48.7%; Pred. No. 0.004;  
XX Matches 133; Conservative 0; Mismatches 140; Indels 0; Gaps 0;

Qy 953 CGATGAGAGTGTGTGTGATCATAGTAGGGGCTCAATGCTTTGACATAGCTTTGGA 1012  
Db 4409 CGTTAGAGATTAGTTGGTTATGTGATTAATTCGTAGTTTACGATATTTTATATAC 4468  
Qy 1013 GAGTGTGATGATGAGATGACTGATGATCGAATATAGTCACATGATGATGATGAG 1012  
Db 4469 GAGTTTAAAGAAATTTTATTTTAAATATTTTATTTGAAATTTTGTATTTAG 4528  
Qy 1073 ACTATTTGTTGTGTGTGTGATGATTAAGGATGATTTGTCAGTTGATTTTGAATTT 1132  
Db 4529 TAAATTTTATTTTATTTTATTTTATTTGATGATTTTATTTTATTTTATTTT 4588  
Qy 1133 CACTTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1192  
Db 4589 TTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTT 4648  
Qy 1193 GTTGACGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1225  
Db 4649 TTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTT 4681

# RESULT 22

ABL33728  
ID ABL33728 standard; DNA; 17721 BP.

AC ABL33728;  
XX

DT 26-MAR-2002 (first entry)  
XX

DE Human immune system associated gene SEQ ID NO: 1701.  
XX

XX Human; immune system disease; cytosine methylation; antiasthmatic;  
XX antiarteriosclerotic; antianaemic; cytosolic; noctropic;  
XX neuroproliferative; anti-HIV; anticonvulsant; ophthalmological;  
XX antirheumatic; antiarthritic; antidiabetic; antipsoriatic;  
XX antineoplastic; cancer; eye disease; arteriosclerosis; anaemia;  
XX acute myeloid leukaemia; Alzheimer's disease; AIDS; epilepsy;  
XX neurofibromatosis; rheumatoid arthritis; psoriasis; bowel disease;  
XX gene; ds.  
XX  
XX Homo sapiens.  
XX  
XX WO200200928-A2.  
XX  
XX 03-JAN-2002.  
XX  
XX 02-JUL-2001; 2001WO-EP07537.  
XX  
XX 30-JUN-2000; 2000DE-1032529.  
XX 01-SEP-2000; 2000DE-1043826.  
XX  
XX (EPIG-) EPIGENOMICS AG.  
XX  
XX Olek A, Piepenbrock C, Berlin K;  
XX  
XX WPI; 2002-130909/17.  
XX

PT Nucleic acid comprising fragment of chemically modified gene, useful  
XX for diagnosis and treatment of diseases associated with abnormal  
XX cytosine methylation -  
XX  
XX Claim 1; SEQ ID NO 1701; 32pp + Sequence Listing; German.  
XX

XX The present invention provides a number of human immune system associated  
XX genes which are modified by the methylation of cytosines. The sequences  
XX can be used in the diagnosis and treatment of immune system disorders,  
XX including eye diseases such as retinopathy, neovascular glaucoma and  
XX macular degeneration, arteriosclerosis, anaemia, cancer, acute myeloid  
XX leukaemia, Alzheimer's disease, AIDS, epilepsy, neurofibromatosis,  
XX rheumatoid arthritis, psoriasis and inflammatory/ulcerative bowel  
XX diseases. The present sequence is a gene of the invention.  
XX

XX Sequence 17721 BP; 4634 A; 328 C; 4081 G; 8678 T; 0 other;







QY 1102 GGAATATGTCAGGATGATATTTGCAAAATCATTGCTGTTTGTTCCTGTT 1161  
 DB 4636 AGAGATGTTGTTTATTTTATATGATGATGTTTGTGAGGTTTATTTAGTGT 4695  
 QY 1162 GTTTGTTCTTTCTCATCTGTAAGCTGTGTTGTAAGTGTGTTGCTGTT 1221  
 DB 4696 GTTAAATTTGAGTTTAAATTTTATTTTATTAATTTAGTTTGGTATGAG 4755  
 QY 1222 GTTTCACACAGAGGTGAGGACTA 1247  
 DB 4756 TTTATTTAGTACGTGAGAAATGATTA 4781

## RESULT 26

ABL54307  
 ID ABL54307 standard; DNA; 7990 BP.

AC ABL54307;

DT 29-JUL-2002 (first entry)

DE Chemically treated apoptosis gene #4.

KM Apoptosis; HIV; Bloom syndrome; cardiopathy;  
 KM neurodegenerative disorder; Herpes simplex virus; renal ischaemia;  
 KM amyotrophic lateral sclerosis; cancer; ds.

OS Unidentified.

PN WO20017164-A2.

PD 18-OCT-2001.

PE 06-APR-2001; 2001WO-EP03969.

PR 06-APR-2000; 2000DE-1019058.

PR 07-APR-2000; 2000DE-1019173.

PR 30-JUN-2000; 2000DE-1032529.

PR 01-SEP-2000; 2000DE-1043826.

PA (EPIG-) EPIGENOMICS AG.

PI Olek A, Piepenbrock C, Berlin K;

PD WPI; 2002-017444/02.

PT Chemically modified sequences of genes associated with apoptosis are  
 PT useful to determine methylation patterns of genomic DNA samples for  
 PT diagnosis of associated diseases such as cancer -

PS Claim 1; Seq ID #7; 24bp; English.

CC This invention relates to chemically pre-treated DNA of genes  
 CC associated with apoptosis. The nucleic acids are used to allocate  
 CC patients for specific therapy for HIV infection, Bloom syndrome,  
 CC cardiopathy, aging, neurodegenerative disorders, Herpes simplex virus  
 CC infection, renal ischaemia, amyotrophic lateral sclerosis, solid tumours  
 CC and cancers. This nucleotide sequence represents a chemically  
 CC treated apoptosis gene. Even SEQ ID numbers are the complementary  
 CC DNA strands to the odd SEQ ID numbers. The sequence data for this  
 CC patent is not represented in the printed specification but is based on  
 CC information supplied by the European patent office.

XX Sequence 7990 BP; 2223 A; 102 C; 1567 G; 4098 T; 0 other;

XX Query Match 3.3%; Score 48.4; DB 24; Length 7990;

XX Best Local Similarity 51.9%; Pred. No. 0.0043;

QY 1016 TGGTGAATGAGAGCTTTCATCGTCAATATAGTGCACATGATGATGAGACT 1075  
 DB 557 TAGAGAAAGTTTAAATTTATTTAGTATATTTATTTATTTATTTTAAAA 616

QY 1076 ATGTGTTGCTGAGTGAAGGCTGATATGTCAGGATGATTTTGCAAATTCAC 1135  
 DB 617 ATTTTATTTATTTTATTTTATTTATTTATTTTATTTTATTTTATTTATTTAT 676  
 QY 1136 TTGTGTTGTTTGTTCCTGTTTGTTCCTGTTTGTTCCTGTTTGTTCCTGTT 1195  
 DB 677 TTGTTTATTTATTTATTTTATTTTATTTTATTTTATTTTATTTTATTTAGAT 736  
 QY 1196 GTACTGTAAGTGTGTTGCTGTTTGT 1225  
 DB 737 TTTATTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTT 766

## RESULT 27

ABL32158  
 ID ABL32158 standard; DNA; 7990 BP.

AC ABL32158;

DT 26-MAR-2002 (first entry)

DE Human immune system associated gene SEQ ID NO: 131.

KM Human; immune system disease; cytosine methylation; antiasthmatic;  
 KM antiarteriosclerotic; antihaemic; cytosolic; noctropic;  
 KM neuroprotective; anti-HIV; anticonvulsant; ophthalmological;  
 KM antirheumatic; antiarthritic; antidiabetic; antiporiatic;  
 KM antiinflammatory; cancer; eye disease; arteriosclerosis; anaemia;  
 KM acute myeloid leukaemia; Alzheimer's disease; AIDS; epilepsy;  
 KM neurofibromatosis; rheumatoid arthritis; psoriasis; bowel disease;  
 KM gene; ds.

OS Homo sapiens.

PN WO200200928-A2.

PD 03-JAN-2002.

PE 02-JUL-2001; 2001WO-EP07537.

PR 30-JUN-2000; 2000DE-1032529.

PR 01-SEP-2000; 2000DE-1043826.

PA (EPIG-) EPIGENOMICS AG.

PI Olek A, Piepenbrock C, Berlin K;

PD WPI; 2002-130909/17.

PT Nucleic acid comprising fragment of chemically modified gene, useful  
 PT for diagnosis and treatment of diseases associated with abnormal  
 PT cytosine methylation -

PS Claim 1; SEQ ID NO 131; 32bp + Sequence Listing; German.

CC The present invention provides a number of human immune system associated  
 CC genes which are modified by the methylation of cytosines. The sequences  
 CC can be used in the diagnosis and treatment of immune system disorders,  
 CC including eye diseases such as retinopathy, neovascular glaucoma and  
 CC macular degeneration, arteriosclerosis, anaemia, cancer, acute myeloid  
 CC leukaemia, Alzheimer's disease, AIDS, epilepsy, neurofibromatosis,  
 CC rheumatoid arthritis, psoriasis and inflammatory/ulcerative bowel  
 CC diseases. The present sequence is a gene of the invention.

XX Sequence 7990 BP; 2223 A; 102 C; 1567 G; 4098 T; 0 other;

XX Query Match 3.3%; Score 48.4; DB 24; Length 7990;

XX Best Local Similarity 51.9%; Pred. No. 0.0043;

QY 1016 TGGTGAATGAGAGCTTTCATCGTCAATATAGTGCACATGATGATGAGACT 1075  
 DB 557 TAGAGAAAGTTTAAATTTATTTAGTATATTTATTTATTTTAAAA 616

QY	1076	ATTGTTGTCGGAGTACGTAAGGGTCGATTTGTCGCACTGATATTTTGCAAATTAC	1135
Db	617	ATTTTTTGGTATTTTGGTTAAATTTAATTTATTTTTTTTTTAATTTTAAAT	676
QY	1136	TTGCGTGTGTTGTTGTCGCGTGTGTTTGTCTTTCATCTGACTGTGTT	1195
Db	677	TGTGTTTAATAATTTGTTTTTTTGGTAAATTTTTTTTAAATTTAATTTAGAT	736
QY	1196	GTACTGACTGTGTGTGGTGTGTTGTT	1225
Db	737	TTTAATATTTTAAATTTAATTTAATTTT	766

RESULT 26	
ABL34014	
ID	ABL34014 standard; DNA; 6182 BP.
XX	
AC	ABL34014;
XX	
DT	26-MAR-2002 (first entry)
XX	
DE	Human immune system associated gene SEQ ID NO: 1987.

Human, immune system disease, cytosine methylation; antiastrumatic; antiartherosclerotic; antianemic; cytosolic; nootropic; neuroprotective; anti-HIV; anticonvulsant; ophthalmological; antirheumatic; antiarthritic; antidiabetic; antisorptic; antineoplastic; cancer; eye disease; arteriosclerosis; anaemia; acute myeloid leukaemia; Alzheimer's disease; AIDS; epilepsy; neurofibromatosis; rheumatoid arthritis; psoriasis; bowel disease; gene; ds.

OS Homo sapiens.  
XX  
FN WO200200928-A2.  
XX  
PD 03-JAN-2002.  
XX  
PF 02-JUL-2001; 2001WO-EP07537.  
XX  
PR 30-JUN-2000; 2000DE-1032529.  
PR 01-SEP-2000; 2000DE-1043826.

AA (EPiG-) EPIGENOMICS AG.  
PA  
XX  
PI Olek A, Piepenbrock C, Berlin K;  
XX  
DR WPI; 2002-130909/17.

PS Claim 1, SEQ ID NO 1987, 35p + Sequence Listing; German.

CC The present invention provides a number of human immune system associated genes which are modified by the methylation of cytosines. The sequences can be used in the diagnosis and treatment of immune system disorders including eye diseases such as retinopathy, neovascular glaucoma or macular degeneration, arteriosclerosis, anemia, cancer, acute myeloid leukaemia, Alzheimer's disease, AIDS, epilepsy, neurofibromatosis or rheumatoid arthritis, psoriasis and inflammatory/alimentary bowel diseases. The present sequence is a gene of the invention.

Sequence 6182 BP; 1929 A; 37 C; 1260 G; 2956 T; 0 other;

Query Match	3.3%;	Score 48.2;	DB 24;	Length 6182;
Best Local Similarity	47.5%;	Pred. No. 0.0044;		
Matches 143; Conservative	0;	Mismatches 158;	Indels 0;	Gaps 0;

Qy 933 TCATGAATATCTATGCAACGAGTGTGTTGTATACATGTAAGTAGGGGTCATG

Db	5388	TTATTAACGAATTTTAAAGAAAGATGTAAATTCGTTAATGATTATTTGTTAGAGTTGAAAT	5447
Qy	933	TGTTTGAGTATGTTGTTTGAGAGCTGATGAGATGAGAGTGACTGATCGTGAATATAGT	10522
Db	5448	TTTTTTAGTCGTTTTTTTTTGAGAGGATGTTTTTTTTTGTCGGGTACGTTTTTTTATTAATAT	55070
Qy	1053	GCAATGATGATGATGATGACATATTTGTTGTGTGGGTAGTGTAAAGGGTGCATATGTC	11122
Db	5508	ATTAATGCTTTTGGGTTTGCAATTTTGTGGAAGAAATATTTGATTTCTGCTTTTTTTTT	5567
Qy	1113	CAGTGAATTTTGCAAATTCACCTGTGTGTTTTGTTTGTTCGCTGTGTTTTCTT	1172
Db	5568	TGGGTATGATGATGATGATTTGCAATTAATTTTGTATTTATTTTATTTATTAATGATCTT	5627
Qy	1173	TCTCATCTGACTGACTGCTGTGTACTGACTGACTGCTGTGTGTTGTTGCACAC	12322
Db	5628	TGTGATTTTGTGTTGGTGTGTTTACTTATTTGTTATTAATAATTTGTTATTTAGGAATAG	5687
Qy	1233	A 1233	
Db	5688	A 5688	

RESULT	29
ABL34147	
ID	ABL34147 standard; DNA; 15667 BP.
XX	
XX	
AC	ABL34147;
XX	
DT	26-MAR-2002 (first entry)
XX	
DE	Human immune system associated gene SEQ ID NO: 2120.

Human, immune system disease; cytosine methylation; antiasthmatic; antiarteriosclerotic; antiamebic; cytosolic; neurotropic; neuropsychotic; anti-HIV; antiinfectant; ophthalmologic; antineumatic; antiarthritic; antibacterial; antipsoriatic; antiinflammatory; cancer; eye disease; arteriosclerosis; anaemia; acute myeloid leukaemia; Alzheimer's disease; AIDS; epilepsy; neurofibromatosis; rheumatoid arthritis; psoriasis; bowel disease; gene; ds.

AA	
OS	Homo sapiens.
XX	
XX	
PN	WO200200928-A2.
XX	
XX	
PD	03-JAN-2002.
XX	
PF	02-JUL-2001, 2001WO-EP07537.
XX	
PR	30-JUN-2000; 2000DE-1032529.
PR	01-SEP-2000; 2000DE-1043826.

AA (EPiG-) EPIGENOMICS AG.  
PA  
XX  
PI Olek A, Piepenbrock C, Berlin K;  
XX  
DR WPI, 2002-130909/17.

PT Nucleic acid comprising fragment of chemically modified gene, useful  
PT for diagnosis and treatment of diseases associated with abnormal  
PT cytosine methylation -  
XX  
PS Claim 1; SEQ ID NO 2120; 32pp + Sequence Listing; German.

The present invention provides a number of human immune system associated genes which are modified by the methylation of cytosines. The sequences can be used in the diagnosis and treatment of immune system disorders, including eye diseases such as retinopathy, neovascular glaucoma and macular degeneration, arteriosclerosis, anaemia, cancer, acute myeloid leukaemia, Alzheimer's disease, AIDS, epilepsy, neurofibromatosis, rheumatoid arthritis, psoriasis and inflammatory/ulcerative bowel diseases. The present sequence is a gene of the invention.

XX Sequence 15667 BP; 4104 A; 186 C; 3634 G; 7743 T; 0 other;  
SQ  
Query Match 3.3%; Score 48.2; DB 24; Length 15667;  
Best Local Similarity 48.7%; Pred. No. 0.0069;  
Matches 131; Conservative 0; Mismatches 138; Indels 0; Gaps 0;  
QY 955 ATGGAAGTGTGTGTGTACATGAGGAGGTCATGTGTGTGAGTATGTGTGAGA 1014  
DB 12230 ATGGAAGATATTGAAATGATGCTTGAAGAAATTTAAATTTATTTATTAATTTATGCGTT 12289  
QY 1015 GTGTGAGATGAGAGACTGCTGATCGTCAATATAGTCACATGTGATTGATGTGAC 1074  
DB 12290 ATTTGTGATTTGATTTGATTTGATTTGATTTGATTTGATTTGATTTGATTTGATTT 12349  
QY 1075 TATTTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1134  
DB 12350 TTTTGGAGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 12409  
QY 1135 CTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1194  
DB 12410 TTTTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 12469  
QY 1195 TGTACTGATCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1223  
DB 12470 TTTATTTTGAGAGGATATAGTATGATGT 12498

RESULT 30  
ABL33990  
ID ABL33990 standard; DNA; 6132 BP.  
XX  
AC ABL33990;  
XX  
DT 26-MAR-2002 (first entry)  
XX  
DE Human immune system associated gene SEQ ID NO: 1963.  
XX  
KW Human; immune system disease; cytosine methylation; antiasthmatic;  
KW antidiabetic; antiaemic; cytosine; neotropic;  
KW antineoplastic; anti-HIV; anticonvulsant; ophthalmological;  
KW antirheumatic; antidiabetic; antidiabetic; antidiabetic;  
KW antineoplastic; cancer; eye disease; arteriosclerosis; anaemia;  
KW acute myeloid leukaemia; Alzheimer's disease; AIDS; epilepsy;  
KW neurofibromatosis; rheumatoid arthritis; psoriasis; bowel disease;  
KW gene; ds.  
XX  
OS Homo sapiens.  
XX  
OS WO200200928-A2.  
XX  
PD 03-JAN-2002.  
XX  
PF 02-JUN-2001; 2001WO-EP07537.  
XX  
PR 30-JUN-2000; 2000DE-1032529.  
XX  
PR 01-SEP-2000; 2000DE-1043826.  
XX  
PA (EPIC-) EPIDENOMICS AG.  
XX  
PI Olek A, Piepenbrock C, Berlin K;  
XX  
DR WPI; 2002-130909/17.  
XX  
PT Nucleic acid comprising fragment of chemically modified gene, useful  
XX for diagnosis and treatment of diseases associated with abnormal  
XX cytosine methylation -  
XX  
PS Claim 1; SEQ ID NO 1963; 32pp + Sequence Listing; German.  
CC The present invention provides a number of human immune system associated  
CC genes which are modified by the methylation of cytosines. The sequences  
CC can be used in the diagnosis and treatment of immune system disorders,

CC including eye diseases such as retinopathy, neovascular glaucoma and  
CC macular degeneration, arteriosclerosis, anaemia, cancer, acute myeloid  
CC leukaemia, Alzheimer's disease, AIDS, epilepsy, neurofibromatosis,  
CC rheumatoid arthritis, psoriasis and inflammatory/ulcerative bowel  
CC diseases. The present sequence is a gene of the invention.  
XX  
SQ Sequence 6132 BP; 1579 A; 115 C; 1717 G; 2721 T; 0 other;  
Query Match 3.2%; Score 48; DB 24; Length 6132;  
Best Local Similarity 48.9%; Pred. No. 0.005;  
Matches 129; Conservative 0; Mismatches 135; Indels 0; Gaps 0;  
QY 960 AGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1019  
DB 2907 AGTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTT 2966  
QY 1020 GAGATGAGAGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1079  
DB 2967 GTGTGAGAGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 3026  
QY 1080 TGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1139  
DB 3027 TTAAGGATGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 3086  
QY 1140 GTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1199  
DB 3087 TTTTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 3146  
QY 1200 TGTACTGATCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1223  
DB 3147 TTTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 3170

RESULT 31  
ABK31536  
ID ABK31536 standard; DNA; 11046 BP.  
XX  
AC ABK31536;  
XX  
DT 23-APR-2002 (first entry)  
XX  
DE Signal transduction associated gene modified DNA #190.  
XX  
KW Human; signal transduction associated gene; cytosine methylation state;  
KW Cpg island; signal transduction associated disease; solid tumour; cancer;  
KW antitumour; cytosine; mutant; ds.  
XX  
OS Homo sapiens.  
XX  
OS Synthetic.  
XX  
OS WO200200926-A2.  
XX  
PD 03-JAN-2002.  
XX  
PF 29-JUN-2001; 2001WO-EP07472.  
XX  
PR 30-JUN-2000; 2000DE-1032529.  
XX  
PR 01-SEP-2000; 2000DE-1043826.  
XX  
PA (EPIC-) EPIDENOMICS AG.  
XX  
PI Olek A, Piepenbrock C, Berlin K;  
XX  
DR WPI; 2002-147896/19.  
XX  
PT Oligonucleotide for diagnosis and therapy of diseases associated with  
XX signal transduction e.g. cancer, comprises chemically modified genomic  
XX sequences of genes associated with signal transduction -  
XX  
PS Claim 1; SEQ ID NO 379; 24pp; English.  
CC The present invention relates to chemically modified DNA sequences of  
CC signal transduction associated genes. The DNA sequences are chemically



ABK28222	ID	ABK28222	standard; DNA; 11394 BP.
XX	AC	ABK28222;	
XX	DT	23-APR-2002	(first entry)
XX	DE	DNA transcription associated complementary genomic DNA #48.	
XX	KW	DNA transcription associated gene; peptide nucleic acid; PNA-oligomer;	
KW	PNA; cytosine methylation state; SNP; retroviral infection; gene; ds;		
KW	single nucleotide polymorphism; adenosine deaminase deficiency; cancer;		
KW	viral infection; Sezary syndrome; haematological disorder; tuberculosis;		
KW	immunological disorder; Werner syndrome; developmental disorder;		
KW	psoriasis; Kieger's syndrome; neurological disorder; erythropoiesis;		
KW	neurodegenerative disorder; Waardenburg syndrome; Niemann-Pick disease;		
KW	myelodysplastic syndrome; myocardial infarction; hypertension; arthritis;		
KW	angiogenesis; congenital heart disease; HDR syndrome; gene therapy;		
KW	polyglutamine disorder; solid tumour.		
XX	XX		
XX	OS	Unidentified.	
XX	WO	WO200192565-A2.	
XX	XX		
XX	PD	06-DEC-2001.	
XX	PF	06-APR-2001; 2001WO-EP03973.	
XX	PR	06-APR-2000; 2000DE-1019058.	
PR	07-APR-2000; 2000DE-1019173.		
PR	30-JUN-2000; 2000DE-1032529.		
PR	01-SEP-2000; 2000DE-1043826.		
XX	PA	(EPIG-) EPIGENOMICS AG.	
XX	PI	Olek A, Piepenbrock C, Berlin K;	
XX	DR	WPI; 2002-090046/12.	
XX	PT	New nucleic acids or oligomers, useful for diagnosing or treating	
PT	diseases associated with DNA transcription, e.g. immunological		
PT	disorders, Werner syndrome, psoriasis, myocardial infarction, solid		
PT	tumours or cancer		
XX	PS	Claim 1; SEQ ID No 96; 32pp; English.	
XX	CC	The invention relates to a nucleic acid, which comprises a segment of the	
CC	CC	chemically pretreated DNA of genes associated with DNA transcription from	
CC	one of 346 sequences, and an oligomer, in particular an oligonucleotide		
CC	or peptide nucleic acid (PNA)-oligomer that hybridises to or is identical		
CC	to the chemically pretreated DNA of genes associated with DNA		
CC	transcription. The set of oligomer probes are useful for detecting the		
CC	cytosine methylation state and/or single nucleotide polymorphisms (SNPs)		
CC	in a chemically pretreated genomic DNA. The nucleic acids are useful for		
CC	diagnosing or treating diseases associated with DNA transcription		
CC	(particularly with the methylation status), e.g. adenosine deaminase		
CC	deficiency, viral infection, retroviral infection, Sezary syndrome,		
CC	haematological disorders, immunological disorders, Werner syndrome,		
CC	tuberculosis, developmental disorders, psoriasis, Kieger's syndrome,		
CC	neurological disorders, neurodegenerative disorders, Waardenburg		
CC	syndrome, Niemann-Pick disease, myelodysplastic syndrome, myocardial		
CC	infarction, hypertension, angiogenesis, erythropoiesis, congenital heart		
CC	disease, HDR syndrome, arthritis, polyglutamine disorders, solid tumours		
CC	or cancer. Sequences ABK28127-ABK28472 represent DNA transcription		
CC	associated genomic DNA molecules of the invention.		
CC	Note: The sequence data for this patent did not form part of the printed		
CC	specification but was obtained in electronic format directly from the		
CC	European Patent Office.		
XX	XX		
XX	Sequence	11394 BP; 3027 A; 200 C; 2257 G; 5910 T; 0 other;	
XX	Query Match	3.2%; Score 47.2; DB 24; Length 11394;	
XX	Best Local Similarity	53.9%; Freq. No. 0.011;	

	Matches	97;	Conservative	0;	Mismatches	83;	Indels	0;	Gaps	
OY	1046	ATATGTCACATGTCATTGTATGTGCAGCTATTGTTGTGGTGATGAACGGTGCAT	1105							
DB	7702	ATAATGATGATGATGATGTTGTTGTGATGAGATATTTAATGCTGAGTTTTTTTTGTTT	7761							
OY	1106	ATTGGCCAGTGGATTAATTTTGGAAATTCATCGATGATGTTTGTTCGCTGATGTT	1165							
DB	7762	TTTAGTTTTTAAATTAATTTAGTTAAATAATTTTTTTTTTTTTTTTTTTTTTTTTT	7821							
OY	1166	TGTTCTTTCTCATCTGACTGACTGACGTTGTTGACTGTACTGTTGTGTTGTTGTT	1225							
DB	7822	TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT	7881							
<b>RESULT 34</b>										
ID	ABLJ33404									
XX	ABLJ33404 standard; DNA; 16033 BP.									
AC	ABLJ33404;									
XX										
DT	26-MAR-2002 (first entry)									
XX										
DE	Human immune system associated gene SEQ ID NO: 1377.									
XX										
KW	Human; immune system disease; cytosine methylation; antiasthmatic;									
KW	antiartherosclerotic; antianaemic; cytostatic; noctropic;									
KW	neuroprotective; anti-HIV; anticonvulsant; ophthalmological;									
KW	antiinflammatory; cancer; eye disease; arteriosclerosis; anaemia;									
KW	acute myeloid leukaemia; Alzheimer's disease; AIDS; epilepsy;									
KW	neurofibromatosis; rheumatoid arthritis; psoriasis; bowel disease;									
gene; ds.										
OS	Homo sapiens.									
PN	WO200200928-A2.									
PD	03-JAN-2002.									
PF	02-JUL-2001; 2001WO-EP07537.									
PR	30-JUN-2000; 2000DE-1032529.									
PR	01-SEP-2000; 2000DE-1043826.									
PA	(EPIG-) EPIDEMIOICS AG.									
PI	Olek A, Piepenbrock C, Berlin K;									
DR	WPI; 2002-130909/17.									
PT	Nucleic acid comprising fragment of chemically modified gene, useful									
PT	for diagnosis and treatment of diseases associated with abnormal									
PT	cytosine methylation -									
PS	Claim 1; SEQ ID NO 1377; 32pp + Sequence listing; German.									
XX										
CC	The present invention provides a number of human immune system associated									
CC	genes which are modified by the methylation of cytosines. The sequences									
CC	can be used in the diagnosis and treatment of immune system disorders,									
CC	including eye diseases such as retinopathy, neovascular glaucoma and									
CC	molecular degeneration, arteriosclerosis, anaemia, cancer, acute myeloid									
CC	leukemia, Alzheimer's disease, AIDS, epilepsy, neurofibromatosis,									
CC	rheumatoid arthritis, psoriasis and inflammatory/ulcerative bowel									
CC	diseases. The present sequence is a gene of the invention.									
SO	Sequence 16033 BP; 3610 A; 352 C; 4561 G; 7510 T; 0 other;									
Query Match	3.2%; Score 47.2; DB 24; Length 16033;									
Best Local Similarity	48.5%; Pred. No. 0.014;									
Matches	130; Conservative 0; Mismatches 138; Indels 0; Gaps 0;									
OY	958 GAAGTGTGTGTGATCATGAAGTACGGGCTCAATGTGTTAGATATGTGTTGGAGACTG	1017								





PA (UYEC-) UNIV EAST CAROLINA.  
XX  
XX NYce JW;  
XX WPI; 2000-205971/18.  
XX  
XX New antisense oligonucleotides useful for treating e.g. pulmonary  
PT vasoconstriction, inflammation, allergies, asthma, hypertension,  
PT bronchitis, emphysema, respiratory distress syndrome, ischemia or  
XX cancers -  
XX  
XX Disclosure; Page 732-736; 1343pp; English.  
XX  
XX The present invention describes a new composition comprising an  
CC antisense oligonucleotide (ON) with low adenosine (up to 15%), which  
CC targets nucleic acids involved in bronchoconstriction, allergies, and/or  
CC inflammation. The ON can have antiinflammatory, antiallergic,  
CC antiasthmatic, cyostatic and analgesic activities. The compositions are  
CC useful for the treatment of diseases associated with inflammation,  
CC impaired airways, including lung disease and diseases whose secondary  
CC effects afflict the lungs of a subject. They can be used for treating  
CC e.g. ischemic conditions, pulmonary vasoconstriction, allergies,  
CC asthma, impeded respiration, respiratory distress syndrome, pain, cystic  
CC fibrosis, pulmonary hypertension, emphysema, chronic obstructive  
CC pulmonary disease (COPD), and cancers such as leukaemias, lymphomas,  
CC carcinomas, and cancers which may metastasize to the lungs, including  
CC breast and prostate cancer. The reduction of the adenosine content of  
CC the ONs reduces side effects. The A-containing ONs break down with the  
CC release of deoxyadenosine which activates adenosine receptors causing  
CC bronchoconstriction and inflammation. AAA3213 to AAA5312 represent the  
CC nucleotide sequences given in the sequence listing from the present  
CC invention, which correspond to SEQ ID NO:1 to 2815, and then the last  
CC 185 sequences are also called SEQ ID NO:1 to 185, but the sequences  
CC differ from the previously named sequences. SEQ ID NO:11 to 1680  
CC (AAA3233 to AAA3392) are specifically claimed ONs from the present  
CC invention. N.B. Sequences given in the disclosure of the present  
CC invention do not match up with their corresponding SEQ ID NO: sequences  
CC given in the sequence listing.  
XX  
SQ Sequence 17073 BP; 3514 A; 4234 C; 5045 G; 4280 T; 0 other;  
Query Match 3.2%; Score 47.2; DB 21; Length 17073;  
Best Local Similarity 50.9%; Pred. No. 0.014;  
Matches 112; Conservative 0; Mismatches 108; Indels 0; Gaps 0;  
QY 673 TGTGAGTATTTCTGTCGTCTGTATCTGTTTGTGTGTCTGTCTACTGAT 732  
DB 13072 TGTCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 13131  
QY 733 CCGGTTGTTGATGGTCAGACCTTGATATTAACAAGCTGTTGGATGACTTTGTAC 792  
DB 13112 TATGTGTGTGTATGTTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 13191  
QY 793 AGTTGAGTAGCAGAGTATTCATCTCCGTCATTTGTTGTGTGAGGTGAGTGAC 852  
DB 13192 TGTCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 13251  
QY 853 TGTGAGGTGAGTTGAGGTGATGATGATGATGATGATGATGATGATGATG 892  
DB 13252 TGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 13291  
RESULT 37  
AAF20999  
ID AAF20999 standard; DNA; 19883 BP.  
XX  
XX AAF20999;  
XX  
XX 14-MAR-2001 (first entry)  
DT  
XX Human low adenosine antisense oligonucleotide related sequence #2566.  
DE  
XX Low adenosine antisense oligonucleotide; phosphorothioate; allergy;

KW human; airway disorder; bronchoconstriction; lung inflammation;  
KW surfactant depletion; respiratory; bronchodilator; antiinflammatory;  
KW immunosuppressive; antiasthmatic; analgesic; hypotensive; cyostatic;  
KW respiratory obstruction; pulmonary vasoconstriction; impeded respiration;  
KW surfactant hypoproduction; pulmonary vasoconstriction; asthma; RDS;  
KW respiratory distress syndrome; pain; cystic fibrosis; allergic rhinitis;  
KW pulmonary hypertension; emphysema; pulmonary transplantation rejection;  
KW chronic obstructive pulmonary disease; pulmonary infection; bronchitis;  
KW cancer; ss.  
XX  
OS Homo sapiens.  
XX  
XX WO200062736-A2.  
XX  
XX 26-OCT-2000.  
XX  
XX 24-MAR-2000; 2000WO-US08020.  
XX  
XX 06-APR-1999; 99US-0127958.  
XX  
XX (UYEC-) UNIV EAST CAROLINA.  
PA (NYCE/) NYCE J W.  
XX  
XX NYce JW;  
XX WPI; 2000-679539/66.  
XX  
XX Low adenosine (A) content antisense oligonucleotides which do not  
PT trigger adenosine receptors during metabolism, useful e.g. for treating  
PT cancers and respiratory obstructions -  
XX  
XX Disclosure; Page 808-813; 1592pp; English.  
XX  
XX The present invention describes low adenosine (A) content antisense  
CC oligonucleotides and compositions (I) comprising them. In the antisense  
CC oligonucleotides the A is replaced by a 'Universal' or alternative base.  
CC (I) can have respiratory, bronchodilator, antiinflammatory, analgesic,  
CC immunosuppressive, antiasthmatic, hypotensive and cyostatic activities.  
CC The antisense oligonucleotides and (I) can be used to down-regulate the  
CC expression and or activity of target polypeptides associated with  
CC lung/respiratory disorders and malignancies, such as stimulating and  
CC activating peptide factors and transmitters, transcription factors,  
CC immunoglobulins and antibodies, antibody receptors, cytokines and  
CC chemokines, endogenously produced specific and non-specific enzymes,  
CC binding proteins, adhesion molecules and their receptors, cytokine and  
CC chemokine receptors, adenosine receptors, bradykinin receptors, central  
CC nervous system (CNS) and peripheral nervous and non-nervous system  
CC receptors, CNS and peripheral nervous and non-nervous system peptide  
CC transmitters, defensins, growth factors, vasoactive peptides and  
CC receptors, binding proteins and malignancy associated proteins. The  
CC antisense oligonucleotides may be used in this way to treat disorders  
CC including respiratory obstruction (especially pulmonary obstruction  
CC and/or bronchoconstriction) and/or lung inflammation, allergy (ies)  
CC and/or surfactant hypoproduction which are associated with a disease or  
CC condition selected from pulmonary vasoconstriction, inflammation,  
CC allergic asthma, impeded respiration, respiratory distress syndrome  
CC (RDS), pain, cystic fibrosis (CF), allergic rhinitis (AR), pulmonary  
CC hypertension, emphysema, chronic obstructive pulmonary disease (COPD),  
CC pulmonary transplantation rejection, pulmonary infections, bronchitis,  
CC and/or cancer. AAF18434 to AAF21543 represent human polynucleotide  
CC fragments and antisense oligonucleotides used in the exemplification of  
CC the present invention.  
XX  
SQ Sequence 19883 BP; 4085 A; 4959 C; 5895 G; 4944 T; 0 other;  
Query Match 3.2%; Score 47.2; DB 21; Length 19883;  
Best Local Similarity 50.9%; Pred. No. 0.015;  
Matches 112; Conservative 0; Mismatches 108; Indels 0; Gaps 0;  
QY 673 TGTGAGTATTTCTGTCGTCTGTATCTGTTTGTGTGTCTGTCTACTGAT 732  
DB 15882 TGTCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 15941

Qy 733 CCGGTTGTTATTTGGTCAGACCTTGATATACACTGTTGGATGACTTTGTGAC 792  
|||  
Db 15942 TATGTCGTCTGATGTTTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTA 16001  
Qy 733 AGTTAGTAGCAGAGTGAATTCATCTCGCTATTTGTGTGTGTGTGTGTGTGTGTGTA 852  
Db 16002 TGTCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTG 16061  
Oy 853 TGTGAGTGATGTTGAGTGATTTGGATGGGATGGAATGTG 892  
Db 16062 TGTGTTCCGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTG 16101  
RESULT 38  
AAA34877  
ID AAA34877 standard; DNA; 19943 BP.  
AC AAA34877;  
XX  
DT 28-JUL-2000 (first entry)  
XX  
DE Human adenosine receptor related polynucleotide SEQ ID NO:2566.  
XX  
KW Human; adenosine receptor; low adenosine antisense oligonucleotide;  
KW phosphorochiolate; impaired respiration; inflammation; allergy;  
KW allergic disease; bronchoconstriction; inhibitor; antiinflammatory;  
KW antiallergic; antiasthmatic; cyostatic; analgesic; impaired airway;  
KW lung disease; ischaemic condition; pulmonary vasoconstriction; asthma;  
KW respiratory distress syndrome; pain; cystic fibrosis; emphysema;  
KW pulmonary hypertension; chronic obstructive pulmonary disease; COPD;  
KW cancer; leukaemia; lymphoma; carcinoma; metastasis; ss.  
XX  
OS Homo sapiens.  
XX  
FN WO200009525-A2.  
XX  
PD 24-FEB-2000.  
XX  
PF 03-AUG-1999; 99WO-US17712.  
XX  
PR 03-AUG-1998; 98US-0095212.  
XX  
PA (UYEC-) UNITV EAST CAROLINA.  
XX  
PI Nyce JW;  
XX  
DR WPI; 2000-205971/18.  
XX  
PT New antisense oligonucleotides useful for treating e.g. pulmonary  
PT vasoconstriction, inflammation, allergies, asthma, hypertension,  
PT bronchitis, emphysema, respiratory distress syndrome, ischemia or  
PT cancers -  
XX  
PS Disclosure; Page 736-741; 1343pp; English.  
XX  
CC The present invention describes a new composition comprising an  
CC antisense oligonucleotide (ON) with low adenosine (up to 15%), which  
CC targets nucleic acids involved in bronchoconstriction, allergies, and/or  
CC inflammation. The ON can have antiinflammatory, antiallergic,  
CC antiasthmatic, cyostatic and analgesic activities. The compositions are  
CC useful for the treatment of diseases associated with inflammation,  
CC impaired airways, including lung disease and diseases whose secondary  
CC effects afflict the lungs of a subject. They can be used for treating  
CC e.g. ischaemic conditions, pulmonary vasoconstriction, allergies,  
CC asthma, impeded respiration, respiratory distress syndrome, pain, cystic  
CC fibrosis, pulmonary hypertension, emphysema, chronic obstructive  
CC pulmonary disease (COPD), and cancers such as leukaemias, lymphomas,  
CC carcinomas, and cancers which may metastasise to the lungs, including  
CC breast and prostate cancer. The reduction of the adenosine content of  
CC the ONs reduces side effects. The A-containing ONs break down with the  
CC release of deoxyadenosine which activates adenosine receptors causing  
CC bronchoconstriction and inflammation. AAA32313 to AAA3312 represent the  
CC nucleotide sequences given in the sequence listing from the present

CC invention, which correspond to SEQ ID NO:1 to 2815, and then the last  
CC 185 sequences are also called SEQ ID NO:1 to 185, but the sequences  
CC differ from the previously named sequences. SEQ ID NO:11 to 1680  
CC (AAA32323 to AAA33992) are specifically claimed ONs from the present  
CC invention. N.B. Sequences given in the disclosure of the present  
CC invention do not match up with their corresponding SEQ ID NO: sequences  
CC given in the sequence listing.  
XX  
SQ Sequence 19943 BP; 4098 A; 4972 C; 5921 G; 4952 T; 0 other;  
Query Match 3.2%; Score 47.2; DB 21; Length 19943;  
Best Local Similarity 50.9%; Pred. No. 0.015; Mismatches 108; Indels 0; Gaps 0;  
Matches 112; Conservative 0; Mismatches 108; Indels 0; Gaps 0;  
Oy 673 TGTCTGTCAGTTATCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTG 732  
|||  
Db 15882 TGTCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTG 15941  
Oy 733 CCGGTTGTTATTTGGTCAGACCTTGATATACACTGTTGGATGACTTTGTGAC 792  
|||  
Db 15942 TATGTCGTCTGATGTTTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTA 16001  
Oy 793 AGTTAGTAGCAGAGTGAATTCATCTCGCTATTTGTGTGTGTGTGTGTGTGTGTGTGTA 852  
Db 16002 TGTCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTG 16061  
Oy 853 TGTGAGTGATGTTGAGTGATTTGGATGGGATGGAATGTG 892  
Db 16062 TGTGTTCCGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTG 16101  
RESULT 39  
ABL49329  
ID ABL49329 standard; DNA; 5666 BP.  
AC ABL49329;  
XX  
DT 01-MAY-2002 (first entry)  
XX  
DE Human polynucleotide associated with DNA replication SEQ ID NO 29.  
XX  
KW Human; cyostatic; neuroprotective; nootropic; immunostimulant;  
KW gene therapy; gene regulation; DNA replication; CNRPB; DNA2U; ATR; CHD1L;  
KW ERCC3; SNRPB1; RAD50; LIG2; cytosine methylation; Ataxia telangiectasia;  
KW ATR-X; Bloom's syndrome; tumour; cancer; methylation; gene; ds.  
XX  
OS Homo sapiens.  
XX  
FN WO200177377-A2.  
XX  
PD 18-OCT-2001.  
XX  
PF 06-APR-2001; 2001WO-EP03971.  
XX  
PR 06-APR-2000; 2000DE-1019058.  
XX  
PR 07-APR-2000; 2000DE-1019173.  
XX  
PR 30-JUN-2000; 2000DE-1032529.  
XX  
PR 01-SEP-2000; 2000DE-1043826.  
XX  
PA (EPIG-) EPIGENOMICS AG.  
XX  
PI Olek A, Piepenbrock C, Berlin K;  
XX  
DR WPI; 2002-017471/02.  
XX  
CC New nucleic acid sequences from chemically modified genes associated  
CC with DNA replication, useful for analysing cytosine methylations for  
CC diagnosis and therapy of diseases e.g. Ataxia telangiectasia -  
XX  
PS Claim 1; SEQ ID NO 29; 23pp + Sequence Listing; English.  
XX  
CC The invention relates to nucleic acid sequences comprising at least 18  
CC bases of a chemically pretreated gene associated with gene regulation,



**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.4 p5\_4578  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 8, 2003, 06:49:04 ; Search time 61 Seconds

(without alignments)  
7425.600 Million cell updates/sec

Title: US-10-020-441-1

Perfect score: 1477  
Sequence: 1 gtgtcaacctgttcgtacacg.....ttgtacgtccaatattga 1477

Scoring table: IDENTITY NUC  
Gapop 10.0 , Gapext 1.0

Searched: 441362 seqs, 153338381 residues

Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents, NA:\*

1: /cgn2\_6/ptodata/1/ina/5A.COMB.seq:\*  
2: /cgn2\_6/ptodata/1/ina/5B.COMB.seq:\*  
3: /cgn2\_6/ptodata/1/ina/6A.COMB.seq:\*  
4: /cgn2\_6/ptodata/1/ina/6B.COMB.seq:\*  
5: /cgn2\_6/ptodata/1/ina/PTOS.COMB.seq:\*  
6: /cgn2\_6/ptodata/1/ina/backfiles1.seq:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	52.4	3.5	7218	1	US-08-232-463-14
2	41.8	2.8	17425	4	US-09-511-625B-5
3	41.4	2.8	19557	5	PCT-US92-06300-1
4	40.4	2.7	3796	2	US-08-762-308-11
5	40.4	2.7	3796	4	US-09-844-634-10
6	40.4	2.7	3813	2	US-08-650-000-3
7	40.4	2.7	3813	6	5395760-3
8	40.2	2.7	700	4	US-09-236-097-11
9	40.2	2.7	3481	4	US-08-965-729A-1
10	39.2	2.7	195	4	US-09-395-604A-24
11	39.2	2.7	4718	3	US-08-936-135-9
12	39.2	2.7	4733	3	US-08-936-135-11
13	39.2	2.7	4769	3	US-08-936-135-13
14	39.2	2.7	4784	3	US-08-936-135-15
15	39	2.6	1733	3	US-09-073-569-1
16	38.2	2.6	1813	4	US-09-071-224-3
17	38	2.6	264	4	US-08-223-177A-10
18	38	2.6	2775	1	US-08-148-096A-1
19	37.8	2.6	870	2	US-08-420-629-9
20	37.6	2.5	191	4	US-09-395-604A-21
21	37.6	2.5	215	4	US-09-395-604A-25
22	37.6	2.5	1066	1	US-08-157-101A-4
23	37.6	2.5	17410	1	US-07-841-646-3
24	37.6	2.5	17410	1	US-08-147-023-3
25	37.6	2.5	17410	1	US-08-447-570-3
26	37.6	2.5	17410	2	US-08-449-700-3
27	37.6	2.5	17410	2	US-08-449-699A-3

28	37.6	2.5	17415	3	US-08-486-343A-1	Sequence 1, Appli
29	37.6	2.5	17415	5	PCT-US95-07349-1	Sequence 1, Appli
30	37.6	2.5	72604	4	US-09-268-992-7	Sequence 7, Appli
31	37.6	2.5	72604	4	US-09-657-474-7	Sequence 7, Appli
32	37.2	2.5	2447	2	US-09-014-969-14	Sequence 14, Appli
33	37.2	2.5	2628	4	US-08-143-219-1	Sequence 1, Appli
34	37	2.5	938	4	US-09-078-294-16	Sequence 16, Appli
35	37	2.5	169998	4	US-09-676-610B-24	Sequence 24, Appli
36	36.8	2.5	2246	4	US-09-363-708-3	Sequence 3, Appli
37	36.8	2.5	2634	4	US-09-463-238-3	Sequence 3, Appli
38	36.8	2.5	2671	6	5168051-9	Patent No. 5168051
39	36.8	2.5	6350	2	US-08-385-335A-9	Sequence 9, Appli
40	36.8	2.5	84495	4	US-09-797-906-3	Sequence 3, Appli
41	36.6	2.5	1454	4	US-09-372-422A-19	Sequence 19, Appli
42	36.4	2.5	2502	1	US-08-062-472B-2	Sequence 2, Appli
43	36.2	2.5	790	4	US-09-363-970-4	Sequence 4, Appli
44	36.2	2.5	1851	2	US-08-418-657D-20	Sequence 20, Appli
45	36.2	2.5	1851	4	US-09-135-080-5	Sequence 5, Appli

## ALIGNMENTS

RESULT 1  
US-08-232-463-14/c  
; Sequence 14, Application US/08232463  
; Patent No. 5670367  
; GENERAL INFORMATION:  
; APPLICANT: DORNER, F.  
; APPLICANT: SCHREIFLINGER, F.  
; APPLICANT: FALKNER, F. G.  
; TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS  
; NUMBER OF SEQUENCES: 52  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Foley & Lardner  
; STREET: 1800 Diagonal Road, Suite 500  
; CITY: Alexandria  
; STATE: VA  
; COUNTRY: USA  
; ZIP: 22313-0299  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/232,463  
; FILING DATE:  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US/07/935,313  
; FILING DATE:  
; APPLICATION NUMBER: EP 91 114 300.6  
; FILING DATE: 26-AUG-1991  
; ATTORNEY/AGENT INFORMATION:  
; NAME: BENT, Stephen A.  
; REGISTRATION NUMBER: 29,768  
; REFERENCE/DOCKET NUMBER: 30472/114 IMTU  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (703) 836-9300  
; TELEFAX: (703) 683-4109  
; TELEX: 899149  
; INFORMATION FOR SEQ ID NO: 14:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 7218 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; IMMEDIATE SOURCE:  
; CLONE: pTZpc-Fls  
; US-08-232-463-14  
Query Match 3.5%; Score 52.4; DB 1; Length 7218;

Best Local Similarity 3.4%; Pred. No. 3e-05;  
Matches 11; Conservative 193; Mismatches 124; Indels 0; Gaps 0;

```
Oy 777 TGGATGACTTTGTGACAGTTCAGTATTCATCTCGGTCACTGTTGGTG 836
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1450 TAGAAGAAATTTGTCACRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1391
Oy 837 AGGTAGAGTACGTATGATGAGTGTGATGATGATGATGATGATGATG 896
    ::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
Db 1390 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1331
Oy 897 GATGGAGTATGAGACACTTGGAGAGAGAACTCATGAATATCTAATG 956
    ::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
Db 1330 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1271
Oy 957 GGAAGTGTCTGTGATCAGAGAGTGGGGGTCAATGTTGATGATGTTG 1016
    ::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
Db 1270 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1211
Oy 1017 GGTGAGATGAGAGTGTGATGTCGATATAGTACATGTGTATGAGACT 1076
    ::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
Db 1210 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1151
Oy 1077 TTGTTGTGGGTAGTGTAAAGGGTGA 1104
    : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1150 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1123
```

RESULT 2  
US-09-511-625B-5  
Sequence 5, Application US/09511625B  
Patent No. 6368828

GENERAL INFORMATION:  
APPLICANT: Larochelle, William J.  
APPLICANT: Patel, Bhaviv H.  
TITLE OF INVENTION: ATTENUATED AND DOMINANT NEGATIVE VARIANT  
TITLE OF INVENTION: CDNA OF STAT6: STAT6B AND STAT6C  
FILE REFERENCE: 14014.0300u1  
CURRENT APPLICATION NUMBER: US/09/511,625B  
CURRENT FILING DATE: 2000-02-23  
PRIOR APPLICATION NUMBER: PCT/US98/17821  
PRIOR FILING DATE: 1998-08-27  
PRIOR APPLICATION NUMBER: 60/070,397  
PRIOR FILING DATE: 1998-01-05  
PRIOR APPLICATION NUMBER: 60/056,075  
PRIOR FILING DATE: 1997-08-27  
NUMBER OF SEQ ID NOS: 70  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 5  
LENGTH: 17425  
TYPE: DNA  
ORGANISM: Artificial Sequence  
FEATURE:  
OTHER INFORMATION: Description of artificial Sequence:/No. 6368828e -  
NAME/KEY: misc feature  
LOCATION: (1)-(17425)  
OTHER INFORMATION: n = a, c, t, c or g  
US-09-511-625B-5

Query Match 2.8%; Score 41.8; DB 4; Length 17425;  
Best Local Similarity 55.0%; Pred. No. 0.055;  
Matches 82; Conservative 0; Mismatches 67; Indels 0; Gaps 0;

```
Oy 652 ATGACTCAGTCTGACAGTATGATGATGATGATGATGATGATGATG 711
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2318 ATGACTCAGAAGTTGAGAACTGTGTGTGTGTGTGTGTGTGTGTGT 2377
Oy 712 GTGTCTGTCTGTACTGATCCGGTTGTTGTTGTTGTTGTTGTTGTTG 771
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2378 GTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 2437
```

```
Oy 772 GTGTTTGATGACTTTGTGACATTCAGT 800
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2438 GTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 2466
```

RESULT 3  
PCT-US92-06300-1/c

Sequence 1, Application PC/TUS9206300  
GENERAL INFORMATION:  
APPLICANT: Hurwitz, David R  
APPLICANT: Nathan, Margaret  
TITLE OF INVENTION: Transgenic Protein Production  
NUMBER OF SEQUENCES: 1  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Rhone-Poulenc Rorer, Inc.  
STREET: 500 Virginia Ave., Bldg. 3A  
CITY: Ft. Washington  
STATE: Pennsylvania  
COUNTRY: USA  
ZIP: 19034  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: PCT/US92/06300  
FILING DATE: 19920730  
CLASSIFICATION: 800  
ATTORNEY/AGENT INFORMATION:  
NAME: Goodman, Rosanne  
REGISTRATION NUMBER: 52,534  
REFERENCE/DOCKET NUMBER: A0856-US  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (215) 962-4130  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1957 base pairs  
TYPE: NUCLEIC ACID  
STRANDEDNESS: double  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
HYPOTHETICAL: NO  
ANTI-SENSE: NO  
PUBLICATION INFORMATION:  
AUTHORS: Mungret, P P  
AUTHORS: Ruffner, D E  
AUTHORS: Kuang, W.-J.  
AUTHORS: Demisson, O E  
AUTHORS: Hawkins, J W  
AUTHORS: Beattie, W G  
AUTHORS: Dugalczyk, A  
TITLE: MOLECULAR STRUCTURE OF THE HUMAN ALBUMIN GENE  
TITLE: IS REVEALED BY NUCLEOTIDE SEQUENCE WITHIN Q11-22  
JOURNAL: J. Biol. Chem.  
VOLUME: 261  
PAGES: 6747-6757  
DATE: 1986  
RELEVANT RESIDUES IN SEQ ID NO: 1: FROM 1 TO 19002  
PCT-US92-06300-1

Query Match 2.8%; Score 41.4; DB 5; Length 19557;  
Best Local Similarity 55.1%; Pred. No. 0.076;  
Matches 81; Conservative 0; Mismatches 66; Indels 0; Gaps 0;

```
Oy 1080 TTGTGTGGTGTGATGAAAGGAGATATGCGCAGTTGATTTTCGAATTC 1139
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 451 TTGAGCTAATGAAATTTATATATATATATATATATATATATATATAT 392
Oy 1140 GTGTTTGTGTTTCCGTGTTGTTTGTCTTTCATCTGATGATGATGATG 1199
```







```
; Sequence 1, Application US/08965729A
; Patent No. 6200751
; GENERAL INFORMATION:
; APPLICANT: Jian-Ming Gu and Charles T. Bamon
; TITLE OF INVENTION: ENDOTHELIAL SPECIFIC EXPRESSION
; TITLE OF INVENTION: REGULATED BY EPCR CONTROL ELEMENTS
; NUMBER OF SEQUENCES: 3
; CORRESPONDENCE ADDRESSES:
; ADDRESSEE: Patrea L. Pabst
; STREET: 2800 One Atlantic Center, 1201 West Peachtree Street
; CITY: Atlanta
; STATE: GA
; COUNTRY: USA
; ZIP: 30309-4530
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/965,729A
; FILING DATE: 07-NOV-1997
; CLASSIFICATION:
; PRIORITY INFORMATION:
; APPLICATION NUMBER: US 60/030,718
; FILING DATE: 08-NOV-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Pabst, Patrea L.
; REGISTRATION NUMBER: 31,284
; REFERENCE/DOCKET NUMBER: OMRF 164 PCT
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 404-873-8795
; TELEFAX: 404-873-8794
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3481 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHEICAL: NO
; ANTI-SENSE: NO
; FEATURE:
; NAME/KEY: misc feature
; OTHER INFORMATION: /note= "Nucleotides 2270 through 2840 are a
; OTHER INFORMATION: large endothelial specific element"; murine
; FEATURE:
; NAME/KEY: misc feature
; OTHER INFORMATION: /note= "Nucleotides 2990 through 3061 are a
; OTHER INFORMATION: serum response element"; murine
; FEATURE:
; NAME/KEY: misc feature
; OTHER INFORMATION: /note= "Nucleotides 3007 through 3014 are a
; OTHER INFORMATION: thrombin responsive element"; murine
; NAME/KEY: misc feature
; OTHER INFORMATION: /note= "Nucleotides 3130 through 3150 are an
; OTHER INFORMATION: endothelial specific element"; murine
; US-08-965-729A-1
Query Match 2.7%; Score 40.2; DB 4; Length 3481;
Best Local Similarity 52.7%; Pred. No. 0.076;
Matches 87; Conservative 0; Mismatches 78; Indels 0; Gaps 0;
```

```
Db 172 ATTAATATTGGCTCCTCAAGATCTGTTTCCCTAAAGAGAGA 128
|||||
RESULT 10
US-09-395-604A-24/C
; Sequence 24, Application US/09395604A
; Patent No. 6458537
; GENERAL INFORMATION:
; APPLICANT: Straub, Rick W.
; APPLICANT: Carrico, Michael M.
; TITLE OF INVENTION: METHODS OF DNA TYPING WITH TANDEM
; TITLE OF INVENTION: REPEATS
; FILE REFERENCE: 062481.0107
; CURRENT APPLICATION NUMBER: US/09/395,604A
; CURRENT FILING DATE: 1999-09-14
; PRIOR FILING DATE: 1996-04-24
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO: 24
; LENGTH: 195
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-395-604A-24
Query Match 2.7%; Score 39.2; DB 4; Length 195;
Best Local Similarity 64.1%; Pred. No. 0.039; Indels 0; Gaps 0;
Matches 59; Conservative 0; Mismatches 33; Indels 0; Gaps 0;
```

```
Qy 638 GGTGATTAACAGCATGACTGCTCAGTCAGATGTCAGTATTCTGTGTCT 697
|||||
Db 194 GGTGACAGAGTAGATGTAAGACTCTCTATATCTATCTATCTGTCTGTCT 135
|||||
Qy 698 GTGTATCTGTTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 729
|||||
Db 134 GTCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 103
|||||
RESULT 11
US-08-936-135-9
; Sequence 9, Application US/08936135
; Patent No. 6054293
; GENERAL INFORMATION:
; APPLICANT: Tessier-Lavigne, Marc
; APPLICANT: He, Zhigang
; TITLE OF INVENTION: Semaphorin Receptors
; NUMBER OF SEQUENCES: 26
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SCIENCE & TECHNOLOGY LAW GROUP
; STREET: 75 DENISE DRIVE
; CITY: HILLSBOROUGH
; STATE: CALIFORNIA
; COUNTRY: USA
; ZIP: 94010
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/936,135
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: OSMAN, RICHARD A
; REGISTRATION NUMBER: 36,627
; REFERENCE/DOCKET NUMBER: UC97-288-2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (650) 343-4341
; TELEFAX: (650) 343-4342
; INFORMATION FOR SEQ ID NO: 9:
```



Sequence 1, Application US/09073569  
 Patent No. 6084088  
 GENERAL INFORMATION:  
 APPLICANT: Sheppard, Paul O.  
 APPLICANT: Grossmann, Angelika  
 TITLE OF INVENTION: NOVEL TUMOR ANTIGENS  
 NUMBER OF SEQUENCES: 15  
 CORRESPONDENCE ADDRESS:  
 ADDRESSEE: Zymogenetics, Inc.  
 STREET: 1201 Baslake Avenue East  
 CITY: Seattle  
 STATE: WA  
 COUNTRY: USA  
 ZIP: 98102  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: Diskette  
 COMPUTER: IBM Compatible  
 OPERATING SYSTEM: DOS  
 SOFTWARE: PASCSEQ for Windows Version 2.0  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/09/073,569  
 FILING DATE:  
 CLASSIFICATION:  
 PRIOR APPLICATION DATA:



```

: INFORMATION FOR SEQ ID NO: 10:
: SEQUENCE CHARACTERISTICS:
:   LENGTH: 264 base pairs
:   TYPE: nucleic acid
:   STRANDEDNESS: double
:   TOPOLOGY: linear
:   MOLECULE TYPE: DNA (genomic)
:   ORIGINAL SOURCE:
:     ORGANISM: Homo sapiens
:     INDIVIDUAL ISOLATE: Caucasian
:     TISSUE TYPE: Blood
:   IMMEDIATE SOURCE:
:     CLONE: M1d22
:   POSITION IN GENOME:
:     CHROMOSOME/SEGMENT: 4
:   FEATURE:
:     NAME/KEY: repeat region
:     LOCATION: 153..193
:     OTHER INFORMATION: /rpt_type="tandem"
:     OTHER INFORMATION: /rpt_family="(dc-da)n.(dg-dt)n"
:     OTHER INFORMATION: /citation=((2))
:   FEATURE:
:     NAME/KEY: misc feature
:     LOCATION: 89..107
:     IDENTIFICATION METHOD: experimental
:     OTHER INFORMATION: /evidence= EXPERIMENTAL
:     OTHER INFORMATION: /standard_name="PCR primer"
:     OTHER INFORMATION: /citation=((1))
:   FEATURE:
:     NAME/KEY: misc feature
:     LOCATION: complement (220..238)
:     IDENTIFICATION METHOD: experimental
:     OTHER INFORMATION: /evidence= EXPERIMENTAL
:     OTHER INFORMATION: /standard_name="PCR primer"
:     OTHER INFORMATION: /citation=((1))
:   FEATURES:
:     NAME/KEY: misc feature
:     LOCATION: 1..264
:     IDENTIFICATION METHOD: experimental
:     OTHER INFORMATION: /evidence= EXPERIMENTAL
:     OTHER INFORMATION: /standard_name="Only one strand sequenced"
:     PUBLICATION INFORMATION:
:       AUTHORS: Weber, J. L.
:       TITLE: Dinucleotide repeat polymorphism at the
:       JOURNAL: Nucleic Acids Res.
:       VOLUME: 18
:       PAGES: 2202-
:       DATE: 1990
:     PUBLICATION INFORMATION:
:       AUTHORS: Weber, James L.
:       TITLE: Abundant Class of Human DNA Polymorphisms
:       TITLE: Which Can Be Typed Using the Polymerase Chain
:       TITLE: Reaction
:       JOURNAL: Am. J. Hum. Genet.
:       VOLUME: 44
:       PAGES: 388-396
:       DATE: 1989
:   US-08-222-177A-10
:
: Query Match          2.6%; Score 38; DB 1; Length 264;
: Best Local Similarity 56.3%; Pred. No. 0.1;
: Matches 71; Conservative 0; Mismatches 55; Indels 0; Gaps 0;

```

```

: QY 772 GTGTTT 777
:   |||
: DB 120 TTATTT 115
:
: RESULT 18
: US-08-149-096A-1
: Sequence 1, Application US/08149096A
: Patent No. 5556956
:
: GENERAL INFORMATION:
:   APPLICANT: ROY, A.K.
:   TITLE OF INVENTION: METHODS AND COMPOSITIONS RELATING TO THE
:   TITLE OF INVENTION: ANDROGEN RECEPTOR GENE AND USES THEREOF
:   NUMBER OF SEQUENCES: 8
:   CORRESPONDENCE ADDRESS:
:     ADDRESS: Arnold, White & Durkee
:     STREET: P.O. Box 4433
:     CITY: Houston
:     STATE: Texas
:     COUNTRY: USA
:   ZIP: 77210
:   COMPUTER READABLE FORM:
:     MEDIUM TYPE: Floppy disk
:     COMPUTER: IBM PC Compatible
:     OPERATING SYSTEM: PC-DOS/MS-DOS
:     SOFTWARE: WordPerfect 5.1
:   CURRENT APPLICATION DATA:
:     APPLICATION NUMBER: US/08/149, 096A
:     FILING DATE: No. 5556956ember 4, 1993
:     CLASSIFICATION: 536
:     ATTORNEY/AGENT INFORMATION:
:       REGISTRATION NUMBER: 33,732
:       REFERENCE/DOCKET NUMBER: UTSK:199\MAX
:       TELECOMMUNICATION INFORMATION:
:         TELEPHONE: (512) 474-7577
:         TELEFAX: (512) 474-7577
:   INFORMATION FOR SEQ ID NO: 1:
:     SEQUENCE CHARACTERISTICS:
:       LENGTH: 2775 base pairs
:       TYPE: nucleic acid
:       STRANDEDNESS: single
:       TOPOLOGY: linear
:     MOLECULE TYPE: oligonucleotide
:   US-08-149-096A-1
:
: Query Match          2.6%; Score 38; DB 1; Length 2775;
: Best Local Similarity 50.5%; Pred. No. 0.3;
: Matches 92; Conservative 0; Mismatches 90; Indels 0; Gaps 0;

```

```

: QY 614 GAAATTTGAAGAAAGTGAAGTTGGTGAATTAACGACATGATCGTCAGTCAGAT 673
: DB 231 GTAAGAAACATTTGGGTGGTGGTGAAGACATTCATGTAACAGATATCC 350
: QY 674 GTCACTAGTATTCTGTGTCGTGTGATCTGTTGGTGTCTGTCTACTGATC 733
: DB 351 TTGGCTTCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 410
: QY 734 CGGTGTGTATTGTGTGACAGCTTGATTAATTAACAAGTGTGGATGATGATGACA 793
: DB 411 GTGTGTCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTCT 470
: QY 794 GT 795
: DB 471 GT 472

```

```

: RESULT 19
: US-08-420-629-9
: Sequence 9, Application US/08420629
: Patent No. 5891627

```

```

; GENERAL INFORMATION:
; APPLICANT: EVANS, GLEN A.
; APPLICANT: SELLER, LUCIA
; APPLICANT: EUBANKS, JAMES H.
; TITLE OF INVENTION: POLYMORPHIC LOCUS
; NUMBER OF SEQUENCES: 9
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SPENSLEY HORN JUBAS & LUBITZ
; STREET: 1880 CENTURY PARK EAST, FIFTH FLOOR
; CITY: LOS ANGELES
; STATE: CALIFORNIA
; COUNTRY: USA
; ZIP: 90067
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/420,629
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/773,099
; FILING DATE: 09-OCT-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: WETHERELL, JR., PH.D., JOHN R.
; REGISTRATION NUMBER: 31,678
; REFERENCE/DOCKET NUMBER: PDI512
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 619-455-5100
; TELEFAX: 619-455-5110
; INFORMATION FOR SEQ ID NO: 9:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 870 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; IMMEDIATE SOURCE:
; CLONE: 4F7
; NAME/KEY: CDS
; LOCATION: 1..870
; US-08-420-629-9

Query Match      2.6%; Score 37.8; DB 2; Length 870;
Best Local Similarity 44.7%; Pred. No. 0.2;
Matches 147; Conservative 0; Mismatches 182; Indels 0; Gaps 0;

QY 682 GTTATCTGAGTGTCTGTGATCTGTTGGTCTGCTGTCTACTGATCCGTTGTT 741
DB 473 GTATGTGTGTGTATGTATGTATGTATGTATGTATGTATGTATGTATGTATGT 532
QY 742 GTATTGTGAGACCTTGTATATACAACTGTGTTGATGACTTTGTGACAGTTGA 801
DB 533 ATATGTATATGTATGTATGTATGTATGTATGTATGTATGTATGTATGTATGT 592
QY 802 GCAGAGTATTTTCATCTCGCTCATTGTGTGTGTGAGAGGTGAGCGTATGTAGG 861
DB 593 ATATGTATATGTATGTATGTATGTATGTATGTATGTATGTATGTATGTATGT 652
QY 862 AGTTGAGTATGTATGTATGTATGTATGTATGTATGTATGTATGTATGTATGT 921
DB 653 ATATGTATATGTATGTATGTATGTATGTATGTATGTATGTATGTATGTATGT 712
QY 922 GGAGAGAGCTCATGAAATATCTATGCAAAAGATGAAAGTGTGTGATCAAGTA 981
DB 713 GTATGTATATGTATGTATGTATGTATGTATGTATGTATGTATGTATGTATGT 772
QY 982 GGGGTCATGTGTGTGAGTATGTGTTG 1010
DB 773 GTATGTATATGTATGTATGTATGTATGTATGTATGTATGTATGTATGTATGT 801
```

```

RESULT 20
US-09-395-604A-21/c
; Sequence 21, Application US/09395604A
; Patent No. 6458537
; GENERAL INFORMATION:
; APPLICANT: Staub, Rick W.
; APPLICANT: Carrico, Michael M.
; TITLE OF INVENTION: METHODS OF DNA TYPING WITH TANDEM
; TITLE OF INVENTION: REPEATS
; FILE REFERENCE: 062481.0107
; CURRENT APPLICATION NUMBER: US/09/395,604A
; CURRENT FILING DATE: 1999-09-14
; PRIOR APPLICATION NUMBER: 08/637,115
; PRIOR FILING DATE: 1996-04-24
; NUMBER OF SEQ ID NOS: 37
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 21
; LENGTH: 191
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-395-604A-21

Query Match      2.5%; Score 37.6; DB 4; Length 191;
Best Local Similarity 63.0%; Pred. No. 0.11;
Matches 58; Conservative 0; Mismatches 34; Indels 0; Gaps 0;

QY 638 GGTGATTAACGACATGACTGCTGCTAGTGCAGATGTCAGTATTCGTGTGCT 697
DB 190 GGTGACAGAGTGAATGAGAGCTCTCTATATCTATCTATCTATCTGTCTGTCT 131
QY 698 GTGTATCTGTTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 729
DB 130 GTCTGTCTGTCTGTCTGTCTGTCTGTCTGTCTGTCTGTCTGTCTGTCTGTCT 99

RESULT 21
US-09-395-604A-25/c
; Sequence 25, Application US/09395604A
; Patent No. 6458537
; GENERAL INFORMATION:
; APPLICANT: Staub, Rick W.
; APPLICANT: Carrico, Michael M.
; TITLE OF INVENTION: METHODS OF DNA TYPING WITH TANDEM
; TITLE OF INVENTION: REPEATS
; FILE REFERENCE: 062481.0107
; CURRENT APPLICATION NUMBER: US/09/395,604A
; CURRENT FILING DATE: 1999-09-14
; PRIOR APPLICATION NUMBER: 08/637,115
; PRIOR FILING DATE: 1996-04-24
; NUMBER OF SEQ ID NOS: 37
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 25
; LENGTH: 215
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-395-604A-25

Query Match      2.5%; Score 37.6; DB 4; Length 215;
Best Local Similarity 63.0%; Pred. No. 0.12;
Matches 58; Conservative 0; Mismatches 34; Indels 0; Gaps 0;

QY 638 GGTGATTAACGACATGACTGCTGCTAGTGCAGATGTCAGTATTCGTGTGCT 697
DB 214 GGTGACAGAGTGAATGAGAGCTCTCTATATCTATCTATCTATCTGTCTGTCT 155
QY 698 GTGTATCTGTTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 729
DB 154 GTCTGTCTGTCTGTCTGTCTGTCTGTCTGTCTGTCTGTCTGTCTGTCTGTCT 123

RESULT 22
```

US-08-157-101A-4/c  
; Sequence 4, Application US/08157101A  
; Patent No. 5808032  
; GENERAL INFORMATION:  
; APPLICANT: KURIHARA, TATSUYA  
; APPLICANT: MATSUKURA, SHIGEKAZU  
; APPLICANT: TSURUOKA, NOBUO  
; APPLICANT: ARIMA, KENJI  
; APPLICANT: NISHIHARA, TATSURO  
; TITLE OF INVENTION: ANTI-HBS ANTIBODY GENES AND EXPRESSION  
; TITLE OF INVENTION: PLASMIDS THEREFOR  
; NUMBER OF SEQUENCES: 9  
; CORRESPONDENCE ADDRESS:  
; ADDRESS: PILLSBURY, MADISON & SUTRO  
; STREET: 1100 NEW YORK AVENUE, N.W.  
; CITY: WASHINGTON  
; STATE: D.C.  
; COUNTRY: USA  
; ZIP: 20005  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentin Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/157,101A  
; FILING DATE: 05-APR-1994  
; CLASSIFICATION: 530  
; ATTORNEY/AGENT INFORMATION:  
; NAME: TITUS, MARLANA K  
; REGISTRATION NUMBER: 35843  
; REFERENCE/DOCKET NUMBER: 9437/204199  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 202-861-3711  
; TELEFAX: 202-822-0944  
; TELEX: 6714627 CUCH  
; INFORMATION FOR SEQ ID NO: 4:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1066 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: DNA (genomic)  
; US-08-157-101A-4  
Query Match 2.5%; Score 37.6; DB 1; Length 1066;  
Best Local Similarity 57.8%; Pred. No. 0.25;  
Matches 67; Conservative 0; Mismatches 49; Indels 0; Gaps 0;  
QY 1113 CAGTGTATATTTTCGAATTCACCTGTGCTGTTGTTGTCGTTGTTGTTCTT 1172  
Db 1050 CAGAGGTAATTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTT 991 /  
QY 1173 TCTCATCTGTAAGTGTGCTGTGTAAGTGTGCTGTGCTGTGCTGTGCTGTGCTGCA 1228  
Db 990 TTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTGA 935  
RESULT 23  
US-07-841-646-3  
; Sequence 3, Application US/07841646  
; Patent No. 5266683  
; GENERAL INFORMATION:  
; APPLICANT: OPPERMANN, HERMANN  
; APPLICANT: OZKAYNAK, ENGIN  
; APPLICANT: KUBERASAMPATH, THANGAVEL  
; APPLICANT: RUEGER, DAVID C.  
; APPLICANT: PANG, ROY H.L.  
; TITLE OF INVENTION: OSTEOGENIC DEVICES  
; NUMBER OF SEQUENCES: 33  
; CORRESPONDENCE ADDRESS:  
; ADDRESS: TESTA, HURWITZ & THIBEAULT  
; STREET: 53 STATE STREET

CITY: BOSTON  
STATE: MASSACHUSETTS  
COUNTRY: U.S.A.  
ZIP: 02109  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/07/841,646  
FILING DATE: 19920221  
CLASSIFICATION: 530  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 810,560  
FILING DATE: 20-DEC-1991  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 827,052  
FILING DATE: 28-JAN-1992  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 660,162  
FILING DATE: 22-FEB-1991  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 621,988  
FILING DATE: 04-DEC-1990  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 621,849  
FILING DATE: 04-DEC-1990  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 616,374  
FILING DATE: 21-NOV-1990  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 600,024  
FILING DATE: 18-OCT-1990  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 599,543  
FILING DATE: 18-OCT-1990  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 579,865  
FILING DATE: 07-SEP-1990  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 569,920  
FILING DATE: 20-AUG-1990  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 483,913  
FILING DATE: 22-FEB-1990  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 422,613  
FILING DATE: 17-OCT-1989  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 315,342  
FILING DATE: 23-FEB-1989  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 232,630  
FILING DATE: 15-AUG-1988  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 179,460  
FILING DATE: 08-APR-1988  
ATTORNEY/AGENT INFORMATION:  
NAME: PITCHER, EDMUND R.  
REGISTRATION NUMBER: 27,829  
REFERENCE/DOCKET NUMBER: CRP-001CP6  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 617/248-7000  
TELEFAX: 617/248-7100  
INFORMATION FOR SEQ ID NO: 3:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 17410 base pairs  
TYPE: NUCLEIC ACID  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
ORIGINAL SOURCE:

ORGANISM: homo sapiens  
FEATURE:  
NAME/KEY: exon  
LOCATION: 3192..3730  
OTHER INFORMATION: //label= EXON-1  
OTHER INFORMATION: //note= "START CODON BEGINS AT POSITION 3313"  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: 10413..10414  
OTHER INFORMATION: //label= GAP-1  
OTHER INFORMATION: //note= "APPROXIMATELY BASES ARE ESTIMATED TO BE MISSING BETWEEN POSITIONS 10413 AND 10414 IN THIS SEQUENCE."  
FEATURE:  
NAME/KEY: exon  
LOCATION: 10696..10891  
OTHER INFORMATION: //label= EXON-2  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: 10960..10961  
OTHER INFORMATION: //label= GAP-2  
OTHER INFORMATION: //note= "APPROXIMATELY BASES ARE ESTIMATED TO BE MISSING BETWEEN POSITION 10960 AND 10961 IN THIS SEQUENCE."  
FEATURE:  
NAME/KEY: exon  
LOCATION: 11059..11211  
OTHER INFORMATION: //label= EXON-3  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: 11351..11352  
OTHER INFORMATION: //label= GAP-3  
OTHER INFORMATION: //note= "APPROXIMATELY BASES ARE ESTIMATED TO BE MISSING BETWEEN POSITIONS 11351 AND 11352 IN THIS SEQUENCE."  
FEATURE:  
NAME/KEY: exon  
LOCATION: 11420..11617  
OTHER INFORMATION: //label= EXON-4  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: 11721..11722  
OTHER INFORMATION: //label= GAP-4  
OTHER INFORMATION: //note= "APPROXIMATELY BASES ARE ESTIMATED TO BE MISSING BETWEEN POSITIONS 11721 AND 11722 IN THIS SEQUENCE."  
FEATURE:  
NAME/KEY: exon  
LOCATION: 13354..13436  
OTHER INFORMATION: //label= EXON-5  
FEATURE:  
NAME/KEY: exon  
LOCATION: 15044..15160  
OTHER INFORMATION: //label= EXON-6  
FEATURE:  
NAME/KEY: exon  
LOCATION: 17245..17410  
OTHER INFORMATION: //label= EXON-7  
US-07-841-646-3

Query Match 2.5%; Score 37.6; DB 1; Length 17410;  
Best Local Similarity 54.3%; Pred. No. 0.92;  
Matches 76; Conservative 0; Mismatches 64; Indels 0; Gaps 0;

QY 1135 CTGTGTTGTTTGTTCCTGTTGTTTCTTCTCATCTGTAAGTTGT 1194  
|||  
Db 4317 CTAAATTTCTTGTGTTCTTCTTGTATTTTCTTTTCTTTTCTTTTC 4376  
|||  
QY 1195 TGTACTGTAAGTTGTTGTTGTTGTTTGTTCACACAGGCGGCGCTAATGGA 1254  
|||  
Db 4377 TTTCTTTCTTTCTTTTCTTTTGTGACGAGGTTTCACTTGTGCGCCAGACTGGA 4436  
|||  
QY 1255 ATGCCGACATGCGACCAATG 1274

Db 4437 GTGCATGCGCGACTCTG 4456  
|||  
RESULT 24  
US-08-147-023-3  
Sequence 3, Application US/08147023  
Patent No. 5468845  
GENERAL INFORMATION:  
APPLICANT: OPPERMAN, HERMANN  
APPLICANT: OZAYNAK, ENGIN  
APPLICANT: KUBERASAMPATH, THANAGAVEL  
APPLICANT: RUEGER, DAVID C.  
APPLICANT: PANG, ROY H.L.  
TITLE OF INVENTION: OSTEOGENIC DEVICES  
NUMBER OF SEQUENCES: 33  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: TESTA, HURWITZ & THIBEAULT  
STREET: 53 STATE STREET  
CITY: BOSTON  
STATE: MASSACHUSETTS  
COUNTRY: U.S.A.  
ZIP: 02109  
COMPUTER READABLE FORM:  
MEDIUM TYPE: floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/147,023  
FILING DATE: 21-FEB-1992  
CLASSIFICATION: 530  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 810,560  
FILING DATE: 20-DEC-1991  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 827,052  
FILING DATE: 28-JAN-1992  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 660,162  
FILING DATE: 22-FEB-1991  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 621,988  
FILING DATE: 04-DEC-1990  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 621,849  
FILING DATE: 04-DEC-1990  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 616,374  
FILING DATE: 21-NOV-1990  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 600,024  
FILING DATE: 18-OCT-1990  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 599,543  
FILING DATE: 18-OCT-1990  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 579,865  
FILING DATE: 07-SEP-1990  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 569,920  
FILING DATE: 20-AUG-1990  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 483,913  
FILING DATE: 22-FEB-1990  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 422,613  
FILING DATE: 17-OCT-1989  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 315,342  
FILING DATE: 23-FEB-1989  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 232,630





APPLICATION NUMBER: US 599,543  
FILING DATE: 18-OCT-1990  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 579,865  
FILING DATE: 07-SEP-1990  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 569,920  
FILING DATE: 20-AUG-1990  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 483,913  
FILING DATE: 22-FEB-1990  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 422,613  
FILING DATE: 17-OCT-1989  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 315,342  
FILING DATE: 23-FEB-1989  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 232,630  
FILING DATE: 15-AUG-1988  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 179,460  
FILING DATE: 08-APR-1988  
ATTORNEY/AGENT INFORMATION:  
NAME: PITCHER, EDMUND R.  
REGISTRATION NUMBER: 27,829  
REFERENCE/DOCKET NUMBER: CRP-001CP6  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 617/248-7000  
TELEFAX: 617/248-7100  
INFORMATION FOR SEQ ID NO: 3:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 17410 base pairs  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
ORGANISM: homo sapiens  
FEATURE:  
NAME/KEY: exon  
LOCATION: 3192..3730  
OTHER INFORMATION: /label= EXON-1  
FEATURE:  
NAME/KEY: misc feature  
LOCATION: 10413..10414  
OTHER INFORMATION: /label= GAP-1  
OTHER INFORMATION: /note= "APPROXIMATELY BASES ARE ESTIMATED TO  
BE MISSING BETWEEN POSITIONS 10413 AND 10414 IN  
OTHER INFORMATION: THIS SEQUENCE."  
FEATURE:  
NAME/KEY: exon  
LOCATION: 10696..10891  
OTHER INFORMATION: /label= EXON-2  
FEATURE:  
NAME/KEY: misc feature  
LOCATION: 10960..10961  
OTHER INFORMATION: /label= GAP-2  
OTHER INFORMATION: /note= "APPROXIMATELY BASES ARE ESTIMATED TO  
BE MISSING BETWEEN POSITIONS 10960 AND 10961 IN  
OTHER INFORMATION: THIS SEQUENCE."  
FEATURE:  
NAME/KEY: exon  
LOCATION: 11059..11211  
OTHER INFORMATION: /label= EXON-3  
FEATURE:  
NAME/KEY: misc feature  
LOCATION: 11351..11352  
OTHER INFORMATION: /label= GAP-3  
OTHER INFORMATION: /note= "APPROXIMATELY BASES ARE ESTIMATED TO  
BE MISSING BETWEEN POSITIONS 11351 AND 11352 IN  
OTHER INFORMATION: THIS SEQUENCE."

FEATURE:  
NAME/KEY: exon  
LOCATION: 11420..11617  
OTHER INFORMATION: /label= EXON-4  
FEATURE:  
NAME/KEY: misc feature  
LOCATION: 11721..11722  
OTHER INFORMATION: /label= GAP-4  
OTHER INFORMATION: /note= "APPROXIMATELY BASES ARE ESTIMATED TO  
BE MISSING BETWEEN POSITIONS 11721 AND 11722 IN  
OTHER INFORMATION: THIS SEQUENCE."  
FEATURE:  
NAME/KEY: exon  
LOCATION: 13354..13436  
OTHER INFORMATION: /label= EXON-5  
FEATURE:  
NAME/KEY: exon  
LOCATION: 15044..15160  
OTHER INFORMATION: /label= EXON-6  
FEATURE:  
NAME/KEY: exon  
LOCATION: 17245..17410  
OTHER INFORMATION: /label= EXON-7  
US-08-447-570-3  
Query Match 2.5%; Score 37.6; DB 1; Length 17410;  
Best Local Similarity 54.3%; Pred. No. 0.92;  
Matches 76; Conservative 0; Mismatches 64; Indels 0; Gaps 0;  
QY 1135 CTGTGCTGTTTGTGTTGTCGCTGTTGTTGTTCTTCTGACTGACTGACTGTT 1194  
DB 4317 CTAAATTTCTTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTG 4376  
QY 1195 TGTACGTACGTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTT 1254  
DB 4377 TTTCTTCTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCT 4436  
QY 1255 ATCCGACATGCGACCAATG 1274  
DB 4437 GTGCAATGGCGGATCTCTG 4456  
RESULT 26  
US-08-449-700-3  
; Sequence 3, Application US/08449700  
; Patent No. 5863758  
; GENERAL INFORMATION:  
; APPLICANT: OPPERMAN, HERMANN  
; APPLICANT: OZKAYNAK, ENGIN  
; APPLICANT: KUBERASAMPATH, THANGAVEL  
; APPLICANT: RUEGER, DAVID C.  
; APPLICANT: PANG, ROY H. L.  
; TITLE OF INVENTION: OSTEOGENIC DEVICES  
; NUMBER OF SEQUENCES: 33  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: TESTA, HURWITZ & THIBEAULT  
; STREET: 53 STATE STREET  
; CITY: BOSTON  
; STATE: MASSACHUSETTS  
; COUNTRY: U.S.A.  
; ZIP: 02109  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/449,700  
; FILING DATE: 21-FEB-1992  
; CLASSIFICATION: 530  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 810,560  
; FILING DATE: 20-DEC-1991

PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 827,052  
FILING DATE: 28-JAN-1992  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 660,162  
FILING DATE: 22-FEB-1991  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 621,988  
FILING DATE: 04-DEC-1990  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 621,849  
FILING DATE: 04-DEC-1990  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 616,374  
FILING DATE: 21-NOV-1990  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 600,024  
FILING DATE: 18-OCT-1990  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 599,543  
FILING DATE: 18-OCT-1990  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 579,865  
FILING DATE: 07-SEP-1990  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 569,920  
FILING DATE: 20-AUG-1990  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 483,913  
FILING DATE: 22-FEB-1990  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 422,613  
FILING DATE: 17-OCT-1989  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 315,342  
FILING DATE: 23-FEB-1989  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 232,630  
FILING DATE: 15-AUG-1988  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 179,460  
FILING DATE: 08-APR-1988  
ATTORNEY/AGENT INFORMATION:  
NAME: PITCHER, EDMUND R.  
REGISTRATION NUMBER: 27,829  
REFERENCE/DOCKET NUMBER: CRP-001CR6  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 617/248-7000  
TELEFAX: 617/248-7100  
INFORMATION FOR SEQ ID NO: 3:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 17410 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
ORIGINAL SOURCE:  
ORGANISM: homo sapiens  
FEATURE:  
NAME/KEY: exon  
LOCATION: 3192..3730  
OTHER INFORMATION: /label= EXON-1  
OTHER INFORMATION: /note= "START CODON BEGINS AT POSITION 3313"  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: 10413..10414  
OTHER INFORMATION: /label= GAP-1  
OTHER INFORMATION: /note= "APPROXIMATELY BASES ARE ESTIMATED TO  
OTHER INFORMATION: BE MISSING BETWEEN POSITIONS 10413 AND 10414 IN  
OTHER INFORMATION: THIS SEQUENCE."  
FEATURE:  
NAME/KEY: exon  
LOCATION: 10696..10891

OTHER INFORMATION: /label= EXON-2  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: 10960..10961  
OTHER INFORMATION: /label= GAP-2  
OTHER INFORMATION: /note= "APPROXIMATELY BASES ARE ESTIMATED TO  
OTHER INFORMATION: BE MISSING BETWEEN POSITION 10960 AND 10961 IN  
OTHER INFORMATION: THIS SEQUENCE."  
FEATURE:  
NAME/KEY: exon  
LOCATION: 11059..11211  
OTHER INFORMATION: /label= EXON-3  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: 11351..11352  
OTHER INFORMATION: /label= GAP-3  
OTHER INFORMATION: /note= "APPROXIMATELY BASES ARE ESTIMATED TO  
OTHER INFORMATION: BE MISSING BETWEEN POSITIONS 11351 AND 11352 IN  
OTHER INFORMATION: THIS SEQUENCE."  
FEATURE:  
NAME/KEY: exon  
LOCATION: 11420..11617  
OTHER INFORMATION: /label= EXON-4  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: 11721..11722  
OTHER INFORMATION: /label= GAP-4  
OTHER INFORMATION: /note= "APPROXIMATELY BASES ARE ESTIMATED TO  
OTHER INFORMATION: BE MISSING BETWEEN POSITIONS 11721 AND 11722 IN  
OTHER INFORMATION: THIS SEQUENCE."  
FEATURE:  
NAME/KEY: exon  
LOCATION: 13354..13436  
OTHER INFORMATION: /label= EXON-5  
FEATURE:  
NAME/KEY: exon  
LOCATION: 15044..15160  
OTHER INFORMATION: /label= EXON-6  
FEATURE:  
NAME/KEY: exon  
LOCATION: 17245..17410  
OTHER INFORMATION: /label= EXON-7  
US-08-449-700-3  
Query Match 2.5%; Score 37.6; DB 2; Length 17410;  
Best Local Similarity 54.3%; Pred. No. 0.92; Mismatches 64; Indels 0; Gaps 0;  
Matches 76; Conservative 0;  
Qy 1135 CTTGCTGTTTGTGTTGTCGTTGTTGTTGTTCTTTCATCTGTAAGTTGT 1194  
Db 4317 CTAATTTCTTTGTTTCTTTCTTTGTTTATTTTCTTTCTTTTCTTTTCTTTTC 4376  
Qy 1195 TGTACTGACTGTTGTTGTTGTTGTTGTTGACACAGGCTGAGCACTAATATGA 1254  
Db 4377 TTTTCTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTT 4376  
Qy 1255 ATGCCGACATGGACCATG 1274  
Db 4437 GTGCAATGGCGGATCTCTG 4456  
RESULT 27  
US-08-449-699A-3  
Sequence 3, Application US/08449699A  
Patent No. 5958441  
GENERAL INFORMATION:  
APPLICANT: OPPERMAN, HERMANN  
APPLICANT: OKAYNAK, ENGIN  
APPLICANT: KUBERASAMPATH, THANAVEL  
APPLICANT: RUEBER, DAVID C.  
APPLICANT: PANG, ROY H.L.  
TITLE OF INVENTION: ANTIBODIES TO OSTEOGENIC PROTEINS  
NUMBER OF SEQUENCES: 33

CORRESPONDENCE ADDRESS:  
ADDRESSEE: TESTA, HURWITZ & THIBEAULT  
STREET: 125 HIGH STREET  
CITY: BOSTON  
STATE: MASSACHUSETTS  
COUNTRY: U.S.A.  
ZIP: 02110  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/449,699A  
FILING DATE: 24-MAY-1995  
CLASSIFICATION: 424  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/147,023  
FILING DATE: 01-NOV-1993  
ATTORNEY/AGENT INFORMATION:  
NAME: PITCHER, EDMUND R.  
REGISTRATION NUMBER: 27,829  
REFERENCE/DOCKET NUMBER: STR-001CP6CN  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 617/248-7000  
TELEFAX: 617/248-7100  
INFORMATION FOR SEQ ID NO: 3:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 17410 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
ORIGINAL SOURCE:  
ORGANISM: homo sapiens  
FEATURE:  
NAME/KEY: exon  
LOCATION: 3192..3730  
OTHER INFORMATION: /label= EXON-1  
OTHER INFORMATION: /note= "START CODON BEGINS AT POSITION 3313"  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: 10413..10414  
OTHER INFORMATION: /label= GAP-1  
OTHER INFORMATION: /note= "APPROXIMATELY BASES ARE ESTIMATED TO  
OTHER INFORMATION: BE MISSING BETWEEN POSITIONS 10413 AND 10414 IN  
OTHER INFORMATION: THIS SEQUENCE."  
FEATURE:  
NAME/KEY: exon  
LOCATION: 10696..10891  
OTHER INFORMATION: /label= EXON-2  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: 10960..10961  
OTHER INFORMATION: /label= GAP-2  
OTHER INFORMATION: /note= "APPROXIMATELY BASES ARE ESTIMATED TO  
OTHER INFORMATION: BE MISSING BETWEEN POSITION 10960 AND 10961 IN  
OTHER INFORMATION: THIS SEQUENCE."  
FEATURE:  
NAME/KEY: exon  
LOCATION: 11059..11211  
OTHER INFORMATION: /label= EXON-3  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: 11351..11352  
OTHER INFORMATION: /label= GAP-3  
OTHER INFORMATION: /note= "APPROXIMATELY BASES ARE ESTIMATED TO  
OTHER INFORMATION: BE MISSING BETWEEN POSITIONS 11351 AND 11352 IN  
OTHER INFORMATION: THIS SEQUENCE."  
FEATURE:  
NAME/KEY: exon  
LOCATION: 11420..11617  
OTHER INFORMATION: /label= EXON-4

FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: 11721..11722  
OTHER INFORMATION: /label= GAP-4  
OTHER INFORMATION: /note= "APPROXIMATELY BASES ARE ESTIMATED TO  
OTHER INFORMATION: BE MISSING BETWEEN POSITIONS 11721 AND 11722 IN  
OTHER INFORMATION: THIS SEQUENCE."  
FEATURE:  
NAME/KEY: exon  
LOCATION: 13354..13436  
OTHER INFORMATION: /label= EXON-5  
FEATURE:  
NAME/KEY: exon  
LOCATION: 15044..15160  
OTHER INFORMATION: /label= EXON-6  
FEATURE:  
NAME/KEY: exon  
LOCATION: 17245..17410  
OTHER INFORMATION: /label= EXON-7  
US-08-449-699A-3  
Query Match 2.5%; Score 37.6; DB 2; Length 17410;  
Best Local Similarity 54.3%; Pred. No. 0.92;  
Matches 76; Conservative 0; Mismatches 64; Indels 0; Gaps 0;  
QY 1135 CTTGTGTGTTTGTGTCGCGTGTGTTGTTCTTTCATCTGACTGTACGTGT 1194  
DB 4317 CTAATTTCTTGTGTTCTTCTTCTTGTATTTTCTTTTCTTTTCTTTTTC 4376  
QY 1195 TGTACTGACTGTGTGCGTGTGTTGTTGCACACAGGCTGAGGACTATATGA 1254  
DB 4377 TTTTCTTTCTTTTCTTTTCTTTTGTAGACGAGTTTACTCTTGTCTGCCAGACTGGA 4436  
QY 1255 ATCCGACATGCGACCAATG 1274  
DB 4437 GTCAATGGCGGACTCTCTG 4456  
RESULT 28  
US-08-486-343A-1  
Sequence 1, Application US/08486343A  
Patent No. 6071695  
GENERAL INFORMATION:  
APPLICANT: OZKAYNAK, ENGİN  
TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR MODULATING  
TITLE OF INVENTION: MORPHOGENIC PROTEIN EXPRESSION  
NUMBER OF SEQUENCES: 7  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: PATENT ADMINISTRATOR, CREATIVE BIOMOLECULES  
STREET: 45 SOUTH STREET  
CITY: HOPKINTON  
STATE: MA  
COUNTRY: USA  
ZIP: 07148  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/486,343A  
FILING DATE: 07-JUN-1995  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: PITCHER, Edmund R.  
REGISTRATION NUMBER: 27,829  
REFERENCE/DOCKET NUMBER: CRP-091CP  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (617)-248-7000  
TELEFAX: (617)-248-7100  
INFORMATION FOR SEQ ID NO: 1:













**THIS PAGE BLANK (USPTO)**



Db 241 TCGCTTGTGTTGTTGATGACAGTGGTTGATGATTTATGCTATATCTTGTC 300  
 Qy 301 TTAATTCAGGTTTCAATTTCTCACTGAGAGAAATGGCAACAAGGCAATTCATCAACA 360  
 Db 301 TTAATTCAGGTTTCAATTTCTCACTGAGAGAAATGGCAACAAGGCAATTCATCAACA 360  
 Qy 361 CCGTCTGAGTTAAGGTGACCAAGGATCATGCCCTCTTATGTCGAGACAGAG 420  
 Db 361 CCGTCTGAGTTAAGGTGACCAAGGATCATGCCCTCTTATGTCGAGACAGAG 420  
 Qy 421 AGACCAATGCGACAGACACTCAGTGGATTCATATTGCAATTGTAAATGCTGCTCAATG 480  
 Db 421 AGACCAATGCGACAGACACTCAGTGGATTCATATTGCAATTGTAAATGCTGCTCAATG 480  
 Qy 481 GTCAACTTACAGAGTGGATCATAGAGATCAGTCCGACAGCAATCGGATATCCGCA 540  
 Db 481 GTCAACTTACAGAGTGGATCATAGAGATCAGTCCGACAGCAATCGGATATCCGCA 540  
 Qy 541 CCGGAACTGTTGTTTCAATTTGTTGATGAGAGGATGATACGACCGTATCCGTCA 600  
 Db 541 CCGGAACTGTTGTTTCAATTTGTTGATGAGAGGATGATACGACCGTATCCGTCA 600  
 Qy 601 CGTAAAGATGTTGAAATATTGAAAGAAAGTGAAGTTGTTGATTAACGACATGACTCAG 660  
 Db 601 CGTAAAGATGTTGAAATATTGAAAGAAAGTGAAGTTGTTGATTAACGACATGACTCAG 660  
 Qy 661 TCGTCAAGTGAATGTCAGTCAATTTCTGTCGTGTCGTGTCGTGTCGTGTCGTGTC 720  
 Db 661 TCGTCAAGTGAATGTCAGTCAATTTCTGTCGTGTCGTGTCGTGTCGTGTCGTGTC 720  
 Qy 721 TCGTCACTGATCCGTTGTTGATTTGATGTCAGAGCCTTGATTAATAACAATGTTGGA 780  
 Db 721 TCGTCACTGATCCGTTGTTGATTTGATGTCAGAGCCTTGATTAATAACAATGTTGGA 780  
 Qy 781 TGAATTTGCAAGTTCACTGACAGAGTATTTCCATCTCGCTCATTTGTTGTTGAGGT 840  
 Db 781 TGAATTTGCAAGTTCACTGACAGAGTATTTCCATCTCGCTCATTTGTTGTTGAGGT 840  
 Qy 841 GAGGTGACGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 900  
 Db 841 GAGGTGACGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 900  
 Qy 901 GGAATGATGACCACTTGAGAGAGAGAGAGACTCATGAATAATCTATGCAAAAGATGAA 960  
 Db 901 GGAATGATGACCACTTGAGAGAGAGAGAGACTCATGAATAATCTATGCAAAAGATGAA 960  
 Qy 961 GTGTGTTGTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1020  
 Db 961 GTGTGTTGTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1020  
 Qy 1021 AGATGAGAGTGAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1080  
 Db 1021 AGATGAGAGTGAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1080  
 Qy 1081 TGTGTGAGTGAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1140  
 Db 1081 TGTGTGAGTGAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1140  
 Qy 1141 TGTGTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTG 1200  
 Db 1141 TGTGTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTG 1200  
 Qy 1201 GTACTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTG 1260  
 Db 1201 GTACTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTG 1260  
 Qy 1261 ACATGAGCAATGAGCAATCTATATGTGTAAGAAGAGGTCAAGATTTTCCGACATTAAC 1320  
 Db 1261 ACATGAGCAATGAGCAATCTATATGTGTAAGAAGAGGTCAAGATTTTCCGACATTAAC 1320  
 Qy 1321 CGCTCAGGTGACAGTGTGAGACCTTCTCTCCATCCCTTCAAGGTCCAGTACTCGGTGT 1380  
 Db 1321 CGCTCAGGTGACAGTGTGAGACCTTCTCTCCATCCCTTCAAGGTCCAGTACTCGGTGT 1380

Db 1321 CGCTCAGGTGACAGTGTGAGACCTTCTCTCCATCCCTTCAAGGTCCAGTACTCGGTGT 1380  
 Qy 1381 CGTATCAGATGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTG 1440  
 Db 1381 CGTATCAGATGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTG 1440  
 Qy 1441 GGTAGAAATGTTGAAATTTTGTACGCTCCAAATATTGA 1477  
 Db 1441 GGTAGAAATGTTGAAATTTTGTACGCTCCAAATATTGA 1477  
 RESULT 2  
 US-09-878-574-4316  
 ; Sequence 4316, Application US/09878574  
 ; Patent No. US20020110548A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Byrum, Joseph R.  
 ; APPLICANT: La Rosa, Thomas J.  
 ; APPLICANT: Thompson, Michael D.  
 ; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated with  
 ; FILE REFERENCE: 38-21(15401)B  
 ; CURRENT APPLICATION NUMBER: US/09/878,574  
 ; PRIOR FILING DATE: 2001-12-21  
 ; PRIOR APPLICATION NUMBER: 09/333,535  
 ; NUMBER OF SEQ ID NOS: 15775  
 ; SEQ ID NO 4316  
 ; LENGTH: 639  
 ; TYPE: DNA  
 ; ORGANISM: Glycine max  
 ; FEATURE:  
 ; NAME/KEY: unsure  
 ; LOCATION: (1)..(639)  
 ; OTHER INFORMATION: unsure at all n locations  
 ; OTHER INFORMATION: Clone ID: LIB3028-014-Q1-B1-H11  
 US-09-878-574-4316  
 Query Match 3.5%; Score 51.2; DB 10; Length 639;  
 Best Local Similarity 52.2%; Pred.No. 3.1e-05;  
 Matches 140; Conservative 0; Mismatches 123; Indels 5; Gaps 1;  
 Qy 962 TGTGTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTG 1020  
 Db 174 TTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTG 233  
 Qy 1021 ---AGATGAGAGTGAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1076  
 Db 234 TTTTGTGAGGAGGTTTGTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTG 293  
 Qy 1077 TTTGTTGAGGAGTGAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1136  
 Db 294 TTTGTTGAGTGTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTG 353  
 Qy 1137 TGTGTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTG 1196  
 Db 354 TTTTGTGAGTGTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTG 1253  
 Qy 1197 TACTGACTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTG 1224  
 Db 414 TGTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTG 441  
 RESULT 3  
 US-09-878-574-4306  
 ; Sequence 4306, Application US/09878574  
 ; Patent No. US20020110548A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Byrum, Joseph R.  
 ; APPLICANT: La Rosa, Thomas J.  
 ; APPLICANT: Thompson, Michael D.  
 ; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated with  
 ; FILE REFERENCE: 38-21(15401)B

```
FILE REFERENCE: 38-21(15401)B
CURRENT APPLICATION NUMBER: US/09/878,574
CURRENT FILING DATE: 2001-12-21
PRIOR APPLICATION NUMBER: 09/333,535
PRIOR FILING DATE: 1999-06-14
NUMBER OF SEQ ID NOS: 15775
SEQ ID NO 4306
LENGTH: 1030
TYPE: DNA
ORGANISM: Glycine max
FEATURE:
NAME/KEY: unsure
LOCATION: (1)..(1030)
OTHER INFORMATION: unsure at all n locations
OTHER INFORMATION: Clone ID: LIB3028-014-Q1-B1-D11
US-09-878-574-4306
```

```
Query Match 3.2% Score 47; DB 10; Length 1030;
Best Local Similarity 40.9%; Pred. No. 0.00079;
Matches 228; Conservative 0; Mismatches 326; Indels 4; Gaps 3;
```

```
QY AGTCGACGTGACGATGTCAGTATTCGTGCTGTGTGATCTGTTGTGTCTG 718
DB AGTCGATNTNTNANNNNNNNNTNTNTGATGNTGATGNTGATGNTGATGNT 208
QY TCTGTCTACCTGATCCGCTGTTGTGATTTGTGACAGCTTGATTAATAACA 778
DB TTTTGTNTNNGTNGTGTGTTGTGTTGTGTTGTGTTGTGTTGTGTTGTGTT 268
QY GATGACTTGTGACAGTTCAGTACGAGTGAATTCATCTCGCTCATCTGTTGT 838
DB TTTTGTGCTGTTGTGTTGTGTTGTGTTGTGTTGTGTTGTGTTGTGTTGTG 328
QY GTGAGGTGACGATGATGATGATGATGATGATGATGATGATGATGATGATG 898
DB GTGAGTGTGTTGTGTTGTGTTGTGTTGTGTTGTGTTGTGTTGTGTTGTGTT 388
QY TGGATGATGACGATGATGATGATGATGATGATGATGATGATGATGATGAT 958
DB GTGTTGTGTTGTGTTGTGTTGTGTTGTGTTGTGTTGTGTTGTGTTGTGTT 448
QY AAGTGTGTTGTGATGATGATGATGATGATGATGATGATGATGATGATGAT 1018
DB GAGTGTGTTGTGTTGTGTTGTGTTGTGTTGTGTTGTGTTGTGTTGTGTT 507
QY TGAGATGAGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1078
DB TGTGTTGTGTTGTGTTGTGTTGTGTTGTGTTGTGTTGTGTTGTGTTGTGTT 567
QY GTTGTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1138
DB GTTGTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 627
QY TGTGTTGTGTTGTGTTGTGTTGTGTTGTGTTGTGTTGTGTTGTGTTGTGTT 1196
DB TGTGTTGTGTTGTGTTGTGTTGTGTTGTGTTGTGTTGTGTTGTGTTGTGTT 686
QY TACTGTACTGTGTTGTGTTGTGTTGTGTTGTGTTGTGTTGTGTTGTGTTGT 1214
DB TGTGTTGTGTTGTGTTGTGTTGTGTTGTGTTGTGTTGTGTTGTGTTGTGTT 704
```

## RESULT 4

```
US-09-878-574-4304
Sequence 4304, Application US/09878574
Patent No. US20020110548A1
GENERAL INFORMATION:
APPLICANT: Byrum, Joseph R.
APPLICANT: Thompson, Michael J.
APPLICANT: Thompson, Michael J.
TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated with
FILE REFERENCE: 38-21(15401)B
```

```
CURRENT APPLICATION NUMBER: US/09/878,574
CURRENT FILING DATE: 2001-12-21
PRIOR APPLICATION NUMBER: 09/333,535
PRIOR FILING DATE: 1999-06-14
NUMBER OF SEQ ID NOS: 15775
SEQ ID NO 4304
LENGTH: 793
TYPE: DNA
ORGANISM: Glycine max
FEATURE:
NAME/KEY: unsure
LOCATION: (1)..(793)
OTHER INFORMATION: unsure at all n locations
OTHER INFORMATION: Clone ID: LIB3028-014-Q1-B1-C11
US-09-878-574-4304
```

```
Query Match 3.1% Score 46.4; DB 10; Length 793;
Best Local Similarity 43.4%; Pred. No. 0.001;
Matches 241; Conservative 0; Mismatches 313; Indels 1; Gaps 1;
```

```
QY TGTGACGTGATTCGTGCTGTGATCTGTTGTGCTGTGCTGCTACCTGAT 732
DB TGTGATGTTGTGTTGTGTTGTGTTGTGTTGTGTTGTGTTGTGTTGTGTT 274
QY CCGTGTGTTGATTTGTGACAGCTTGATTAATAACAACCTGTTGTGATGAC 792
DB TGTGTTGTGTTGTGTTGTGTTGTGTTGTGTTGTGTTGTGTTGTGTTGTGTT 334
QY AGTTGATGACAGATGATTCATCTCGCTCATCTGTTGTGTTGTGTTGTGTT 852
DB TGTGTTGTGTTGTGTTGTGTTGTGTTGTGTTGTGTTGTGTTGTGTTGTGTT 394
QY TGTGAGTGTGATGATGATGATGATGATGATGATGATGATGATGATGATG 912
DB TGTGTTGTGTTGTGTTGTGTTGTGTTGTGTTGTGTTGTGTTGTGTTGTGTT 454
QY CCATGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 972
DB TGTGTTGTGTTGTGTTGTGTTGTGTTGTGTTGTGTTGTGTTGTGTTGTGTT 514
QY CATGAGTGTGTTGTGATGATGATGATGATGATGATGATGATGATGATGAT 1031
DB TGTGTTGTGTTGTGTTGTGTTGTGTTGTGTTGTGTTGTGTTGTGTTGTGTT 574
QY GACTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1091
DB GTGTTGTGTTGTGTTGTGTTGTGTTGTGTTGTGTTGTGTTGTGTTGTGTT 634
QY TGTAAAGGTGTGATGATGATGATGATGATGATGATGATGATGATGATGAT 1151
DB TGTGTTGTGTTGTGTTGTGTTGTGTTGTGTTGTGTTGTGTTGTGTTGTGTT 694
QY GTTCCGTTGTGTTGTGTTGTGTTGTGTTGTGTTGTGTTGTGTTGTGTT 1211
DB TGTGTTGTGTTGTGTTGTGTTGTGTTGTGTTGTGTTGTGTTGTGTTGTGTT 754
QY TGTGTTGTGTTGTGTTGTGTTGTGTTGTGTTGTGTTGTGTTGTGTTGTGTT 1226
DB TGTGTTGTGTTGTGTTGTGTTGTGTTGTGTTGTGTTGTGTTGTGTTGTGTT 769
```

## RESULT 5

```
US-09-960-352-9095
Sequence 9095, Application US/09960352
Patent No. US20020137139A1
GENERAL INFORMATION:
APPLICANT: Warren, Wesley C.
APPLICANT: Byatt, John C.
APPLICANT: Mathiasen, Nagappan
TITLE OF INVENTION: NUCLEIC ACID AND OTHER MOLECULES ASSOCIATED WITH LACTATION AND
FILE REFERENCE: 16511.006/37-21(10298)C
```



```

; APPLICANT: Thompson, Michael D.
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated with
; FILE OF INVENTION: Plants
; FILE REFERENCE: 38-21(15401)B
; CURRENT APPLICATION NUMBER: US/09/878,574
; CURRENT FILING DATE: 2001-12-21
; PRIOR APPLICATION NUMBER: 09/333,535
; PRIOR FILING DATE: 1999-06-14
; NUMBER OF SEQ ID NOS: 15775
; SEQ ID NO 4310
; LENGTH: 584
; TYPE: DNA
; ORGANISM: Glycine max
; NAME/KEY: unsure
; LOCATION: (1)..(584)
; OTHER INFORMATION: unsure at all n locations
; OTHER INFORMATION: Clone ID: LIB3028-014-Q1-B1-F10
US-09-878-574-4310

```

```

Query Match          3.0%; Score 44; DB 10; Length 584;
Best Local Similarity 55.1%; Pred. No. 0.0046;
Matches 86; Conservative 0; Mismatches 70; Indels 0; Gaps 0;

```

```

QY 1070 TCGACTATTGTTGTTGGGTAGTAAAGGTGATATGTCAGTATATTTTCGAA 1129
DB 405 TGTATTTTGTGTTGTTGTTTGTGTTTGTGTTTGTGTTTGTGTTTGTGTTTGT 464
QY 1130 ATTCACTTGTGTTTGTGTTTGTGTTTGTGTTTGTGTTTGTGTTTGTGTTTGT 1189
DB 465 TTTTGTGTTTGTGTTTGTGTTTGTGTTTGTGTTTGTGTTTGTGTTTGTGTTTGT 524
QY 1190 GTTGTGTAAGTACTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTT 1225
DB 525 GTTTCGCGTTTGTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTT 560

```

```

RESULT 9
US-09-910-943-4/c
; Sequence 4, Application US/09910943
; Patent No. US20020081610A1
; GENERAL INFORMATION:
; APPLICANT: Hemmati-Briyvanlou, Ali
; TITLE OF INVENTION: Assays and Materials for Embryonic Gene Expression
; FILE REFERENCE: 7529/1G148US1
; CURRENT APPLICATION NUMBER: US/09/910,943
; CURRENT FILING DATE: 2001-07-23
; NUMBER OF SEQ ID NOS: 742
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 4
; LENGTH: 749
; TYPE: DNA
; ORGANISM: Xenopus laevis
; NAME/KEY: misc.feature
; LOCATION: (1)..(749)
; OTHER INFORMATION: n may be a or g or c or t/u
US-09-910-943-4

```

```

Query Match          3.0%; Score 43.8; DB 10; Length 749;
Best Local Similarity 56.6%; Pred. No. 0.0062;
Matches 81; Conservative 0; Mismatches 62; Indels 0; Gaps 0;

```

```

QY 683 TTAATCTGTGTCTGTGTATCTGTTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 742
DB 368 TAAATCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 309
QY 743 TATTGTGAGAGCTTGATTAATACACATGTTTGGATGACTTGTGACAGTGTAG 802
DB 308 TGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTAA 249
QY 803 CAGAGTATTCATCTCGGTCA 825

```

```

DB 248 CAGCAGCAGCTTCTCAGCAGCA 226

```

```

RESULT 10
US-09-844-653-5
; Sequence 5, Application US/09844653
; Publication No. US20030054347A1
; GENERAL INFORMATION:
; APPLICANT: Richards, Julia
; TITLE OF INVENTION: Detecting and Treating Eye Disease
; FILE REFERENCE: UM-06105
; CURRENT APPLICATION NUMBER: US/09/844,653
; CURRENT FILING DATE: 2001-04-27
; NUMBER OF SEQ ID NOS: 173
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 5
; LENGTH: 402850
; TYPE: DNA
; ORGANISM: Homo sapiens
; NAME/KEY: misc.feature
; LOCATION: (46565)..(46565)
; OTHER INFORMATION: The n at this position can be a, c, t, or g.
; NAME/KEY: misc.feature
; LOCATION: (11743)..(118256)
; OTHER INFORMATION: The n at these positions can be a, c, t, or g.
; NAME/KEY: misc.feature
; LOCATION: (118272)..(118736)
; OTHER INFORMATION: The n at these positions can be a, c, t, or g.
; NAME/KEY: misc.feature
; LOCATION: (120172)..(120640)
; OTHER INFORMATION: The n at these positions can be a, c, t, or g.
; NAME/KEY: misc.feature
; LOCATION: (122654)..(122654)
; OTHER INFORMATION: The n at these positions can be a, c, t, or g.
; NAME/KEY: misc.feature
; LOCATION: (203527)..(203963)
; OTHER INFORMATION: The n at these positions can be a, c, t, or g.
; NAME/KEY: misc.feature
; LOCATION: (215340)..(215340)
; OTHER INFORMATION: The n at this position can be a, c, t, or g.
; NAME/KEY: misc.feature
; LOCATION: (220602)..(220650)
; OTHER INFORMATION: The n at these positions can be a, c, t, or g.
; NAME/KEY: misc.feature
; LOCATION: (222415)..(222416)
; OTHER INFORMATION: The n at these positions can be a, c, t, or g.
US-09-844-653-5

```

```

Query Match          2.9%; Score 43.2; DB 9; Length 402850;
Best Local Similarity 55.3%; Pred. No. 0.47;
Matches 104; Conservative 0; Mismatches 83; Indels 1; Gaps 1;

```

```

QY 691 TGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 750
DB 312520 TGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 312579
QY 751 AGAGCCTTGATTAAT-AAACAAGTGTGTTGATGACTTGTGACAGTGTAGAGAG 809
DB 312580 CATGTTTGTGTTTGTGCTTCAATGTTTGTGTTTGTGTTTGTGTTTGTGTT 312639
QY 810 ATTTCATCTCGGTGATTTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 869
DB 312640 CTTTCCTTTTGTGTTTGTGTTTGTGAGTGAATTCATCTGTGACCTGTGG 312699
QY 870 TAGATTGG 877
DB 312700 TGTAAAGG 312707

```

RESULT 11





```
/ NAME/KEY: misc_feature
/ LOCATION: (602466)..(602485)
/ OTHER INFORMATION: n is unidentified a, c, g, or t
/ NAME/KEY: misc_feature
/ LOCATION: (546598)..(547017)
/ OTHER INFORMATION: n is unidentified a, c, g, or t
/ NAME/KEY: misc_feature
/ LOCATION: (494715)..(494814)
/ OTHER INFORMATION: n is unidentified a, c, g, or t
/ NAME/KEY: misc_feature
/ LOCATION: (390586)..(391005)
/ OTHER INFORMATION: n is unidentified a, c, g, or t
/ NAME/KEY: misc_feature
/ LOCATION: (346860)..(346823)
/ OTHER INFORMATION: n is unidentified a, c, g, or t
/ NAME/KEY: misc_feature
/ LOCATION: (317174)..(317193)
/ OTHER INFORMATION: n is unidentified a, c, g, or t
/ NAME/KEY: misc_feature
/ LOCATION: (280353)..(280373)
/ OTHER INFORMATION: n is unidentified a, c, g, or t
/ NAME/KEY: misc_feature
/ LOCATION: (271829)..(271848)
/ OTHER INFORMATION: n is unidentified a, c, g, or t
/ NAME/KEY: misc_feature
/ LOCATION: (183872)..(183891)
/ OTHER INFORMATION: n is unidentified a, c, g, or t
/ NAME/KEY: misc_feature
/ LOCATION: (170625)..(170645)
/ OTHER INFORMATION: n is unidentified a, c, g, or t
/ NAME/KEY: misc_feature
/ LOCATION: (132680)..(132700)
/ OTHER INFORMATION: n is unidentified a, c, g, or t
/ NAME/KEY: misc_feature
/ OTHER INFORMATION: n is a, c, g, or t
/ US-09-771-208-20
```

```
Query Match 2.9%; Score 42.4; DB 9; Length 659158;
Best Local Similarity 55.4%; Pred. No. 1.1;
Matches 82; Conservative 0; Mismatches 66; Indels 0; Gaps 0;
```

```
Qy 1077 TTGTTGTGGTGTAGTGAAGGAGGATATTTGTCAGTTGATTTTGCATTCAC 1136
Db 317280 TTCTCTTGTGGGCTGTGTTTGTGTTGCTTTCCTTTGTTGCTTCTCT 317339
Qy 1137 TGTGTTGTTTGTGTTGTCGTTGTTGTTGTTGTTGTTGTTGTTGTTGTT 1196
Db 317340 TCCCGTGTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTT 317399
Qy 1197 TACTGACTGTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTT 1224
Db 317400 TTCTGTGTTCTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTT 317427
```

```
RESULT 15
US-09-960-352-12328
/ Sequence 12328, Application US/09960352
/ Patent No. US20020137139A1
/ GENERAL INFORMATION:
/ APPLICANT: Warren, Wesley C.
/ APPLICANT: Tao, Nengbing
/ APPLICANT: Byatt, John C.
/ APPLICANT: Mathialagan, Nagappan
/ TITLE OF INVENTION: NUCLEIC ACID AND OTHER MOLECULES ASSOCIATED WITH LACTATION AND
/ TITLE OF INVENTION: MUSCLE AND FAT DEPOSITION
/ FILE REFERENCE: 16511.006/37-21(10298)C
/ CURRENT FILING DATE: 2001-09-24
/ NUMBER OF SEQ ID NOS: 15112
/ SEQ ID NO 12328
/ LENGTH: 395
/ TYPE: DNA
/ ORGANISM: Bos taurus
```

```
/ OTHER INFORMATION: Clone ID: 53-LIB188-014-Q1-E1-F2
/ US-09-960-352-12328
```

```
Query Match 2.8%; Score 41.6; DB 10; Length 395;
Best Local Similarity 52.3%; Pred. No. 0.019;
Matches 92; Conservative 0; Mismatches 84; Indels 0; Gaps 0;
```

```
Qy 1124 TTGCAATTCACCTGTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTT 1183
Db 10 TACAGCTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTT 69
Qy 1184 CTGTAAGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTT 1243
Db 70 TTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTT 129
Qy 1244 ACTATATATGATGCGGACATCGGACCAATGCAATCCATATATGTTGTAACAG 1299
Db 130 TTTATTTGCTGAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 185
```

```
RESULT 16
US-09-960-352-13486
/ Sequence 13486, Application US/09960352
/ Patent No. US20020137139A1
/ GENERAL INFORMATION:
```

```
/ APPLICANT: Warren, Wesley C.
/ APPLICANT: Tao, Nengbing
/ APPLICANT: Byatt, John C.
/ APPLICANT: Mathialagan, Nagappan
/ TITLE OF INVENTION: NUCLEIC ACID AND OTHER MOLECULES ASSOCIATED WITH LACTATION AND
/ TITLE OF INVENTION: MUSCLE AND FAT DEPOSITION
/ FILE REFERENCE: 16511.006/37-21(10298)C
/ CURRENT FILING DATE: 2001-09-24
/ NUMBER OF SEQ ID NOS: 15112
/ SEQ ID NO 13486
/ LENGTH: 403
/ TYPE: DNA
/ ORGANISM: Bos taurus
/ OTHER INFORMATION: Clone ID: 58-BOVMS1-008-Q1-E1-G10
/ US-09-960-352-13486
```

```
Query Match 2.8%; Score 41.6; DB 10; Length 403;
Best Local Similarity 57.8%; Pred. No. 0.02;
Matches 74; Conservative 0; Mismatches 54; Indels 0; Gaps 0;
```

```
Qy 1109 GTCCAGTTGATATTTTGCATTCACCTGTGTTGTTGTTGTTGTTGTTGTTGTT 1168
Db 6 GCGTCATTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTT 65
Qy 1169 TCTTTCTATCTGTAAGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTT 1228
Db 66 TTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTT 125
Qy 1229 CCACAGGG 1236
Db 126 CCACAGGG 133
```

```
RESULT 17
US-09-960-352-1009
/ Sequence 1009, Application US/09960352
/ Patent No. US20020137139A1
/ GENERAL INFORMATION:
/ APPLICANT: Warren, Wesley C.
/ APPLICANT: Tao, Nengbing
/ APPLICANT: Byatt, John C.
/ APPLICANT: Mathialagan, Nagappan
/ TITLE OF INVENTION: NUCLEIC ACID AND OTHER MOLECULES ASSOCIATED WITH LACTATION AND
/ TITLE OF INVENTION: MUSCLE AND FAT DEPOSITION
/ FILE REFERENCE: 16511.006/37-21(10298)C
/ CURRENT FILING DATE: 2001-09-24
```

```
/ NUMBER OF SEQ ID NOS: 15112
/ SEQ ID NO: 1009
/ LENGTH: 439
/ TYPE: DNA
/ ORGANISM: Bos taurus
/ FEATURE:
/ NAME/KEY: unsure
/ LOCATION: (62)
/ OTHER INFORMATION: unsure at all n locations
/ OTHER INFORMATION: Clone ID: 05-LIB3057-015-Q1-K1-B9
US-09-966-352-1009
```

```
Query Match          2.8%; Score 41.6; DB 10; Length 439;
Best Local Similarity 47.3%; Pred. No. 0.021;
Matches 125; Conservative 0; Mismatches 139; Indels 0; Gaps 0;
```

```
QY 962 TGTGTGTGTACATGAGAGGGGTCATGCTTTGAGTAGTGTGAGAGTGTGA 1021
DB 156 TGTGTGTGTGTCTTATTTATTTTCTGTTTTTTTTTTTTTTTTTTTTT 215
QY 1022 GATGAGAGTGAATGATGTCGATATATAGTGACATGATGATGACTATTGTT 1081
DB 216 TATTTTTTTTATTTATTTTCTTTTATTTTATTTTATTTTATTTTATTTT 275
QY 1082 GTGTGGGTAGTAAAGGTGATATGTGCGAGTTGATATTTTGAATTCATGTGT 1141
DB 276 TTTTTTTTTTTTTTTTTTTTTTTTTTGTGGGGGTGTTTTTTTTTTTTTATTT 335
QY 1142 GTTTTGTGTGTCCTGTGTGTTTCTTCTCATCTGTAAGTGTGTTGTTGTTG 1201
DB 336 TTTTCTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTATTT 395
QY 1202 TACTGTGTGTGTTGTTGTTT 1225
DB 396 TATTTATTTCTTATTTTATTTT 419
```

## RESULT 18

```
US-09-964-824A-98/c
/ Sequence 98; Application US/09964824A
/ Patent No. US20020102531A1
/ GENERAL INFORMATION:
/ APPLICANT: Horrigan, Stephen
/ TITLE OF INVENTION: Cancer Gene Determination and Therapeutic Screening Using Signatu
/ FILE REFERENCE: 689290-73
/ CURRENT APPLICATION NUMBER: US/09/964,824A
/ PRIOR FILING DATE: 2001-09-27
/ PRIOR APPLICATION NUMBER: US/60/236,033
/ PRIOR FILING DATE: 2000-09-28
/ PRIOR APPLICATION NUMBER: US/60/236,032
/ PRIOR FILING DATE: 2000-09-28
/ PRIOR APPLICATION NUMBER: US/60/236,028
/ PRIOR FILING DATE: 2000-09-28
/ NUMBER OF SEQ ID NOS: 583
/ SOFTWARE: PatentIn version 3.0
/ SEQ ID NO: 98
/ LENGTH: 10476
/ TYPE: DNA
/ ORGANISM: Homo sapiens
US-09-964-824A-98
```

```
Query Match          2.8%; Score 41.6; DB 10; Length 10476;
Best Local Similarity 64.6%; Pred. No. 0.15;
Matches 62; Conservative 0; Mismatches 34; Indels 0; Gaps 0;
```

```
QY 687 TCTGTGTGTGTGTATCTGTTGTGTGTCTGTCTACCGATCCGTTGTATT 746
DB 3024 TCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 2965
QY 747 GGTCAAGCCTTGATATATACAACTGTGTTGATG 782
DB 2964 GGTATGTGTCTGATGAGAGTCTGTGTGTGTG 2929
```

## RESULT 19

```
US-09-964-824A-552/c
/ Sequence 552; Application US/09964824A
/ Patent No. US20020102531A1
/ GENERAL INFORMATION:
/ APPLICANT: Horrigan, Stephen
/ TITLE OF INVENTION: Cancer Gene Determination and Therapeutic Screening Using Signatu
/ FILE REFERENCE: 689290-73
/ CURRENT APPLICATION NUMBER: US/09/964,824A
/ PRIOR FILING DATE: 2001-09-27
/ PRIOR APPLICATION NUMBER: US/60/236,033
/ PRIOR FILING DATE: 2000-09-28
/ PRIOR APPLICATION NUMBER: US/60/236,032
/ PRIOR FILING DATE: 2000-09-28
/ PRIOR APPLICATION NUMBER: US/60/236,028
/ PRIOR FILING DATE: 2000-09-28
/ NUMBER OF SEQ ID NOS: 583
/ SOFTWARE: PatentIn version 3.0
/ SEQ ID NO: 552
/ LENGTH: 10476
/ TYPE: DNA
/ ORGANISM: Homo sapiens
US-09-964-824A-552
```

```
Query Match          2.8%; Score 41.6; DB 10; Length 10476;
Best Local Similarity 64.6%; Pred. No. 0.15;
Matches 62; Conservative 0; Mismatches 34; Indels 0; Gaps 0;
```

```
QY 687 TCTGTGTGTGTGTATCTGTTGTGTGTCTGTCTACCGATCCGTTGTATT 746
DB 3024 TCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 2965
QY 747 GGTCAAGCCTTGATATATACAACTGTGTTGATG 782
DB 2964 GGTATGTGTCTGATGAGAGTCTGTGTGTGTG 2929
```

## RESULT 20

```
US-09-967-768A-314
/ Sequence 314; Application US/09967768A
/ Patent No. US20020150877A1
/ GENERAL INFORMATION:
/ APPLICANT: Augustus, Meena
/ TITLE OF INVENTION: Cancer Gene Determination and Therapeutic Screening Using Signatu
/ FILE REFERENCE: 689290-72
/ CURRENT APPLICATION NUMBER: US/09/967,768A
/ PRIOR FILING DATE: 2001-09-28
/ PRIOR APPLICATION NUMBER: US/60/236,109
/ PRIOR FILING DATE: 2000-09-28
/ PRIOR APPLICATION NUMBER: US/60/236,034
/ PRIOR FILING DATE: 2000-09-28
/ PRIOR APPLICATION NUMBER: US/60/236,111
/ PRIOR FILING DATE: 2000-09-28
/ NUMBER OF SEQ ID NOS: 325
/ SOFTWARE: PatentIn version 3.0
/ SEQ ID NO: 314
/ LENGTH: 174424
/ TYPE: DNA
/ ORGANISM: Homo sapiens
US-09-967-768A-314
```

```
Query Match          2.8%; Score 41.6; DB 10; Length 174424;
Best Local Similarity 49.5%; Pred. No. 0.86;
Matches 107; Conservative 0; Mismatches 109; Indels 0; Gaps 0;
```

```
QY 678 GTCAATATCTGTGTGTGTGTATCTGTTGTGTGTCTGTCTACCGATCCGTT 737
DB 147515 GTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 147574
```







```

: GENERAL INFORMATION:
: APPLICANT: Byrum, Joseph R.
: APPLICANT: La Rosa, Thomas J.
: APPLICANT: Thompson, Michael D.
: TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated with
: TITLE OF INVENTION: Plants
: FILE REFERENCE: 38-21 (15401) B
: CURRENT APPLICATION NUMBER: US/09/878,574
: CURRENT FILING DATE: 2001-12-21
: PRIOR APPLICATION NUMBER: 09/333,535
: PRIOR FILING DATE: 1999-06-14
: NUMBER OF SEQ ID NOS: 15775
: SEQ ID NO 4299
: LENGTH: 545
: TYPE: DNA
: ORGANISM: Glycine max
: FEATURE:
: NAME/KEY: unsure
: LOCATION: (1)..(545)
: OTHER INFORMATION: unsure at all n locations
: OTHER INFORMATION: Clone ID: LTB3028-013-Q1-B1-H1
ITS-09-878-574-4299

```

Query Match 2.7%; Score 40.6; DB 10; Length 545;  
 Best Local Similarity 54.3%; Pred. No. 0.048;  
 Matches 82; Conservative 0; Mismatches 69; Indels 0; Gaps 0;

Qy 1075 TATGTGTGCTGAGTGTAAAGGTGATATGTGCCAGTGTGATTTTGAATTC 1134  
 Db 147 TTTTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 206  
 Qy 1135 CTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1194  
 Db 207 TTGGTGGGGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 266  
 Qy 1195 TGTACTGTACTGTGTGTGTGTGTGTGTGT 1225  
 Db 267 TTTGTGTGTGTGTGTGTGTGTGTGTGTGT 297

RESULT 35  
 US-09-917-800A-1591/c  
 ; Sequence 1591, Application US/09917800A  
 ; Patent No. US20020119462A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Mendrick, Donna  
 ; APPLICANT: Potter, Mark  
 ; APPLICANT: Johnson, Kory  
 ; APPLICANT: Casale, Arthur  
 ; APPLICANT: Elashoff, Michael  
 ; APPLICANT: Gene Logic, Inc.  
 ; TITLE OF INVENTION: Molecular Toxicology Modeling  
 ; FILE REFERENCE: 44921-5038-US  
 ; CURRENT APPLICATION NUMBER: US/09/917,800A  
 ; CURRENT FILING DATE: 2001-07-31  
 ; PRIOR APPLICATION NUMBER: US 60/222,040  
 ; PRIOR FILING DATE: 2000-07-31  
 ; PRIOR APPLICATION NUMBER: US 60/222,880  
 ; PRIOR FILING DATE: 2000-11-02  
 ; PRIOR APPLICATION NUMBER: US 60/290,029  
 ; PRIOR FILING DATE: 2001-05-11  
 ; PRIOR APPLICATION NUMBER: US 60/290,645  
 ; PRIOR FILING DATE: 2001-05-15  
 ; PRIOR APPLICATION NUMBER: US 60/292,336  
 ; PRIOR FILING DATE: 2001-05-22  
 ; PRIOR APPLICATION NUMBER: US 60/295,798  
 ; PRIOR FILING DATE: 2001-06-06  
 ; PRIOR APPLICATION NUMBER: US 60/297,457  
 ; PRIOR FILING DATE: 2001-06-13  
 ; PRIOR APPLICATION NUMBER: US 60/298,884  
 ; PRIOR FILING DATE: 2001-06-19  
 ; PRIOR APPLICATION NUMBER: US 60/303,459  
 ; PRIOR FILING DATE: 2001-07-09  
 ; NUMBER OF SEQ ID NOS: 1740  
 ; SOFTWARE: PatentIn Ver. 2.1  
 ; SEQ ID NO 1591  
 ; LENGTH: 1545  
 ; TYPE: DNA  
 ; ORGANISM: Rattus norvegicus  
 ; FEATURE:  
 ; OTHER INFORMATION: Genbank Accession No. US20020119462A1 NM\_012977  
 ; US-09-917-800A-1591

Query Match 2.7%; Score 40.6; DB 10; Length 1545;  
 Best Local Similarity 54.3%; Pred. No. 0.091;  
 Matches 82; Conservative 0; Mismatches 69; Indels 0; Gaps 0;  
 Qy 1116 TTGAATTTTGAATTCCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1175  
 Db 1539 TTTTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1480  
 Qy 1176 CATCTGACTGACGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1235  
 Db 1479 TTTTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1420  
 Qy 1236 GTGAGCGACTAATGAATGCCGACATGC 1266

Db 1419 AAAGATTGGGAAAGATATGTGACATTC 1389

RESULT 36  
 US-10-028-072-37/c  
 ; Sequence 37, Application US/10028072  
 ; Publication No. US2003004311A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Baker, Kevin P.  
 ; APPLICANT: Beresini, Maureen  
 ; APPLICANT: DeForge, Laura  
 ; APPLICANT: Desnoyers, Luc  
 ; APPLICANT: Filvaroff, Ellen  
 ; APPLICANT: Gao, Wei-Qiang  
 ; APPLICANT: Gerlisen, Mary E.  
 ; APPLICANT: Goddard, Audrey  
 ; APPLICANT: Godowski, Paul J.  
 ; APPLICANT: Gurney, Austin L.  
 ; APPLICANT: Sherwood, Steven  
 ; APPLICANT: Smith, Victoria  
 ; APPLICANT: Stewart, Timothy A.  
 ; APPLICANT: Tumas, Daniel  
 ; APPLICANT: Watanabe, Colin K  
 ; APPLICANT: Wood, William  
 ; APPLICANT: Zhang  
 ; TITLE OF INVENTION:  
 ; FILE REFERENCE:  
 ; CURRENT APPLICATION NUMBER: US/10/028,072  
 ; CURRENT FILING DATE: 2001-12-19  
 ; PRIOR APPLICATION NUMBER: 60/049911  
 ; PRIOR FILING DATE: 1997-06-18  
 ; PRIOR APPLICATION NUMBER: 60/056974  
 ; PRIOR FILING DATE: 1997-08-26  
 ; PRIOR APPLICATION NUMBER: 60/059113  
 ; PRIOR FILING DATE: 1997-09-17  
 ; PRIOR APPLICATION NUMBER: 60/059115  
 ; PRIOR FILING DATE: 1997-09-17  
 ; PRIOR APPLICATION NUMBER: 60/059117  
 ; PRIOR FILING DATE: 1997-09-17  
 ; PRIOR APPLICATION NUMBER: 60/059122  
 ; PRIOR FILING DATE: 1997-09-17  
 ; PRIOR APPLICATION NUMBER: 60/059184  
 ; PRIOR FILING DATE: 1997-09-17  
 ; PRIOR APPLICATION NUMBER: 60/059263  
 ; PRIOR FILING DATE: 1997-09-18  
 ; PRIOR APPLICATION NUMBER: 60/059352  
 ; PRIOR FILING DATE: 1997-09-19  
 ; PRIOR APPLICATION NUMBER: 60/059588  
 ; PRIOR FILING DATE: 1997-09-19  
 ; PRIOR APPLICATION NUMBER: 60/059836  
 ; PRIOR FILING DATE: 1997-09-24  
 ; PRIOR APPLICATION NUMBER: 60/062250  
 ; PRIOR FILING DATE: 1997-10-17  
 ; PRIOR APPLICATION NUMBER: 60/062285  
 ; PRIOR FILING DATE: 1997-10-17  
 ; PRIOR APPLICATION NUMBER: 60/062287  
 ; PRIOR FILING DATE: 1997-10-17  
 ; PRIOR APPLICATION NUMBER: 60/062814  
 ; PRIOR FILING DATE: 1997-10-24  
 ; PRIOR APPLICATION NUMBER: 60/062816  
 ; PRIOR FILING DATE: 1997-10-24  
 ; PRIOR APPLICATION NUMBER: 60/063045  
 ; PRIOR FILING DATE: 1997-10-24  
 ; PRIOR APPLICATION NUMBER: 60/063082  
 ; PRIOR FILING DATE: 1997-10-31  
 ; PRIOR APPLICATION NUMBER: 60/063127  
 ; PRIOR FILING DATE: 1997-10-24  
 ; PRIOR APPLICATION NUMBER: 60/063327  
 ; PRIOR FILING DATE: 1997-10-27  
 ; PRIOR APPLICATION NUMBER: 60/063329  
 ; PRIOR FILING DATE: 1997-10-27  
 ; PRIOR APPLICATION NUMBER: 60/063550

PRIOR FILING DATE: 1997-10-28  
PRIOR APPLICATION NUMBER: 60/063561  
PRIOR FILING DATE: 1997-10-28  
PRIOR APPLICATION NUMBER: 60/063704  
PRIOR FILING DATE: 1997-10-29  
PRIOR APPLICATION NUMBER: 60/063733  
PRIOR FILING DATE: 1997-10-29  
PRIOR APPLICATION NUMBER: 60/063735  
PRIOR FILING DATE: 1997-10-29  
PRIOR APPLICATION NUMBER: 60/063738  
PRIOR FILING DATE: 1997-10-29  
PRIOR APPLICATION NUMBER: 60/063755  
PRIOR FILING DATE: 1997-10-17  
PRIOR APPLICATION NUMBER: 60/064248  
PRIOR FILING DATE: 1997-11-03  
PRIOR APPLICATION NUMBER: 60/064809  
PRIOR FILING DATE: 1997-11-07  
PRIOR APPLICATION NUMBER: 60/065186  
PRIOR FILING DATE: 1997-11-12  
PRIOR APPLICATION NUMBER: 60/065846  
PRIOR FILING DATE: 1997-11-17  
PRIOR APPLICATION NUMBER: 60/066364  
PRIOR FILING DATE: 1997-11-21  
PRIOR APPLICATION NUMBER: 60/066453  
PRIOR FILING DATE: 1997-11-24  
PRIOR APPLICATION NUMBER: 60/066511  
PRIOR FILING DATE: 1997-11-24  
PRIOR APPLICATION NUMBER: 60/066770  
PRIOR FILING DATE: 1997-11-24  
PRIOR APPLICATION NUMBER: 60/069212  
PRIOR FILING DATE: 1997-12-11  
PRIOR APPLICATION NUMBER: 60/069278  
PRIOR FILING DATE: 1997-12-11  
PRIOR APPLICATION NUMBER: 60/069334  
PRIOR FILING DATE: 1997-12-11  
PRIOR APPLICATION NUMBER: 60/069694  
PRIOR FILING DATE: 1997-12-16  
PRIOR APPLICATION NUMBER: 60/072320  
PRIOR FILING DATE: 1998-01-23  
PRIOR APPLICATION NUMBER: 60/073612  
PRIOR FILING DATE: 1998-02-04  
PRIOR APPLICATION NUMBER: 60/074086  
PRIOR FILING DATE: 1998-02-09  
PRIOR APPLICATION NUMBER: 60/074092  
PRIOR FILING DATE: 1998-02-09  
PRIOR APPLICATION NUMBER: 60/077791  
PRIOR FILING DATE: 1998-03-12  
PRIOR APPLICATION NUMBER: 60/078910  
PRIOR FILING DATE: 1998-03-20  
PRIOR APPLICATION NUMBER: 60/079294  
PRIOR FILING DATE: 1998-03-25  
PRIOR APPLICATION NUMBER: 60/079663  
PRIOR FILING DATE: 1998-02-27  
PRIOR APPLICATION NUMBER: 60/079728  
PRIOR FILING DATE: 1998-03-27  
PRIOR APPLICATION NUMBER: 60/080165  
PRIOR FILING DATE: 1998-03-31  
PRIOR APPLICATION NUMBER: 60/081203  
PRIOR FILING DATE: 1998-04-09  
PRIOR APPLICATION NUMBER: 60/081229  
PRIOR FILING DATE: 1998-04-09  
PRIOR APPLICATION NUMBER: 60/081695  
PRIOR FILING DATE: 1998-04-14  
PRIOR APPLICATION NUMBER: 60/081817  
PRIOR FILING DATE: 1998-04-15  
PRIOR APPLICATION NUMBER: 60/081818  
PRIOR FILING DATE: 1998-04-15  
PRIOR APPLICATION NUMBER: 60/082999  
PRIOR FILING DATE: 1998-04-24  
PRIOR APPLICATION NUMBER: 60/083322  
PRIOR FILING DATE: 1998-04-28  
PRIOR APPLICATION NUMBER: 60/083545  
PRIOR FILING DATE: 1998-04-29

PRIOR APPLICATION NUMBER: 60/084600  
PRIOR FILING DATE: 1998-05-07  
PRIOR APPLICATION NUMBER: 60/084627  
PRIOR FILING DATE: 1998-05-07  
PRIOR APPLICATION NUMBER: 60/084637  
PRIOR FILING DATE: 1998-05-07  
PRIOR APPLICATION NUMBER: 60/085149  
PRIOR FILING DATE: 1998-05-12  
PRIOR APPLICATION NUMBER: 60/085323  
PRIOR FILING DATE: 1998-05-13  
PRIOR APPLICATION NUMBER: 60/085338  
PRIOR FILING DATE: 1998-05-13  
PRIOR APPLICATION NUMBER: 60/085339  
PRIOR FILING DATE: 1998-05-13  
PRIOR APPLICATION NUMBER: 60/085579  
PRIOR FILING DATE: 1998-05-15  
PRIOR APPLICATION NUMBER: 60/085697  
PRIOR FILING DATE: 1998-05-15  
PRIOR APPLICATION NUMBER: 60/085704  
PRIOR FILING DATE: 1998-05-15  
PRIOR APPLICATION NUMBER: 60/086414  
PRIOR FILING DATE: 1998-05-22  
PRIOR APPLICATION NUMBER: 60/086430  
PRIOR FILING DATE: 1998-05-22  
PRIOR APPLICATION NUMBER: 60/087106  
PRIOR FILING DATE: 1998-05-28  
PRIOR APPLICATION NUMBER: 60/088026  
PRIOR FILING DATE: 1998-06-04  
PRIOR APPLICATION NUMBER: 60/088730  
PRIOR FILING DATE: 1998-06-10  
PRIOR APPLICATION NUMBER: 60/088741  
PRIOR FILING DATE: 1998-06-10  
PRIOR APPLICATION NUMBER: 60/088810  
PRIOR FILING DATE: 1998-06-10  
PRIOR APPLICATION NUMBER: 60/088858  
PRIOR FILING DATE: 1998-06-11  
PRIOR APPLICATION NUMBER: 60/089532  
PRIOR FILING DATE: 1998-06-17  
PRIOR APPLICATION NUMBER: 60/089599  
PRIOR FILING DATE: 1998-06-17  
PRIOR APPLICATION NUMBER: 60/089907  
PRIOR FILING DATE: 1998-06-18  
PRIOR APPLICATION NUMBER: 60/089947  
PRIOR FILING DATE: 1998-06-19  
PRIOR APPLICATION NUMBER: 60/090349  
PRIOR FILING DATE: 1998-06-23  
PRIOR APPLICATION NUMBER: 60/090429  
PRIOR FILING DATE: 1998-06-24  
PRIOR APPLICATION NUMBER: 60/090445  
PRIOR FILING DATE: 1998-06-24  
PRIOR APPLICATION NUMBER: 60/090538  
PRIOR FILING DATE: 1998-06-24  
PRIOR APPLICATION NUMBER: 60/090863  
PRIOR FILING DATE: 1998-06-26  
PRIOR APPLICATION NUMBER: 60/091360  
PRIOR FILING DATE: 1998-07-01  
PRIOR APPLICATION NUMBER: 60/091519  
PRIOR FILING DATE: 1998-07-02  
PRIOR APPLICATION NUMBER: 60/091982  
PRIOR FILING DATE: 1998-07-07

Query Match 2.7%; Score 40.6; DB 9; Length 3501;  
Best Local Similarity 50.3%; Pred. No. 0.15;  
Matches 100; Conservative 0; Mismatches 99; Indels 0; Gaps 0;

QY 1116 TGCATTTTCGAATTCATCTGTGTGTTGTTGTCGTTGTTGTTGTTCT 1175  
DB 3497 TTTTGTCTGACGTCGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTT 3438  
QY 1176 CATCTGACGTCGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGACG 1235  
DB 3437 TTTTGTCTGACGTCGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTT 3378



Publication No. US20030022328A1

APPLICANT: Sherwood, Steven

RESULT 39  
US-10-140-470-37/C  
Sequence 37, Application US/10140470  
Publication NO. US20030022331A1  
GENERAL INFORMATION:  
APPLICANT: Baker, Kevin P.  
APPLICANT: Berselin, Maureen  
APPLICANT: DeForge, Laura  
APPLICANT: Desnoyers, Luc  
APPLICANT: Filvaroff, Ellen  
APPLICANT: Gao, Wei-Qiang  
APPLICANT: Gerritsen, Mary E.  
APPLICANT: Goddard, Andrew  
APPLICANT: Godowski, Paul J.  
APPLICANT: Gurney, Auecin L.  
APPLICANT: Sherwood, Steven  
APPLICANT: Smith, Victoria



GenCore version 5.1.4\_p5\_4578  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 8, 2003, 06:44:20 ; Search time 1802 Seconds  
(without alignments)  
13274.547 Million cell updates/sec

Title: US-10-020-441-1  
Perfect score: 1477  
Sequence: 1 gtgtcaacctgtgtcatgacg.....ttgtacgtccaatattga 1477

Scoring table: IDENTITY NUC  
Gapop 10.0 , Gapext 1.0

Searched: 16154066 seqs, 8097743376 residues  
Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :  
EST:  
1: em\_estba:\*  
2: em\_esthum:\*  
3: em\_estin:\*  
4: em\_estmu:\*  
5: em\_estov:\*  
6: em\_estpl:\*  
7: em\_estro:\*  
8: em\_hic:\*  
9: gb\_est1:\*  
10: gb\_est2:\*  
11: gb\_hic:\*  
12: gb\_est3:\*  
13: gb\_est4:\*  
14: gb\_est5:\*  
15: em\_estfun:\*  
16: em\_estom:\*  
17: gb\_gss:\*  
18: em\_gss\_hum:\*  
19: em\_gss\_inv:\*  
20: em\_gss\_pln:\*  
21: em\_gss\_vrt:\*  
22: em\_gss\_fun:\*  
23: em\_gss\_mam:\*  
24: em\_gss\_mus:\*  
25: em\_gss\_other:\*  
26: em\_gss\_pro:\*  
27: em\_gss\_rod:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	75.8	5.1	1101	CNS00LT2	AL078714 Drosophila
2	71.4	4.8	895	CNS0071A	AL066286 Drosophila
3	68.6	4.6	884	CNS006U0	AL065823 Drosophila
4	67.8	4.6	1101	CNS00LO0	AL068607 Drosophila
5	66	4.5	1068	CNS00ETV	AL069846 Drosophila
6	65.4	4.4	960	CNS005F3	AL059925 Drosophila

7	64.8	4.4	1038	17	CNS0108N	AL098657 Drosophila
8	64.6	4.4	922	17	CNS0073W	AL066784 Drosophila
9	62.6	4.2	1101	17	CNS00FXE	AL071370 Drosophila
10	62.4	4.2	1101	17	CNS0100X	AL098379 Drosophila
11	62.4	4.2	1101	17	CNS0172T	AL108707 Drosophila
12	62	4.2	1225	17	CNS0166K	AL106358 Drosophila
13	61.8	4.2	1101	17	CNS04WIP	AL1310426 Tetradodon
14	60.2	4.1	939	17	CNS00CNG	AL059400 Drosophila
15	59.4	4.0	1101	17	CNS0039G	AL063921 Drosophila
16	59.2	4.0	1114	17	A0743348	A0743348 HS 5387 B
17	59	4.0	1101	17	CNS01844	AL108862 Drosophila
18	58.8	4.0	542	17	CNS02Y82	AL219356 Tetradodon
19	58.8	4.0	1101	17	CNS00LIM	AL066473 Drosophila
20	58.6	4.0	909	17	CNS00JTL	AL076720 Drosophila
21	58.4	4.0	1101	17	CNS0106X	AL098595 Drosophila
22	56.4	3.8	1069	17	A0743361	A0743361 HS 5387 B
23	56.4	3.8	1101	17	CNS000D1	AL065414 Drosophila
24	56	3.8	967	13	BM415882	BM415882 OP20964 M
25	55.4	3.8	948	17	CNS003OT	AL064477 Drosophila
26	55.4	3.8	1101	17	CNS0039E	AL063919 Drosophila
27	55.2	3.7	859	17	CNS004XY	AL055406 Drosophila
28	55.2	3.7	993	17	CNS005N6	AL060418 Drosophila
29	55	3.7	942	17	CNS00601	AL065607 Drosophila
30	55	3.7	987	17	CNS017LB	AL108185 Drosophila
31	54.8	3.7	791	17	CNS009KS	AL053801 Drosophila
32	54.6	3.7	1068	17	CNS00J53	AL076666 Drosophila
33	54.4	3.7	1131	17	CNS03AWY	AL235768 Tetradodon
34	54.2	3.7	963	17	AG135074	AG135074 Pan trogl
35	54	3.7	778	17	CNS010UP	AL099451 Drosophila
36	54	3.7	916	17	CNS0126V	AL101185 Drosophila
37	53.8	3.6	860	17	CNS018FL	AL109275 Drosophila
38	53.4	3.6	929	13	BM415054	BM415054 OP20123 M
39	53.2	3.6	1101	17	CNS016MS	AL1065942 Drosophila
40	53.2	3.6	1109	12	BG030246	BG030246 602297544
41	53	3.6	1035	17	CNS00ZD5	AL097523 Drosophila
42	53	3.6	1119	12	BG775904	BG775904 602650307
43	52.6	3.6	833	17	CNS00XRS	AL095254 Arabidops
44	52.6	3.6	833	17	CNS007X3	AL050945 Drosophila
45	52.6	3.6	861	17	CNS0075A	AL066834 Drosophila

ALIGNMENTS

RESULT 1  
LOCUS  
DEFINITION  
ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL

CNS00LT2 1101 bp DNA linear GSS 14-JUN-1999  
Drosophila melanogaster genome survey sequence TERT end of BAC:  
BACR48P19 of RPCI-98 library from Drosophila melanogaster (fruit  
fly), genomic survey sequence.  
AL078714  
AL078714.1 GI:5102004  
GSS.  
Drosophila melanogaster.  
Drosophila melanogaster  
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;  
Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
Ephydroidea; Drosophilidae; Drosophila.  
1 (bases 1 to 1101)  
Genoscope.  
Direct Submission  
Submitted (11-JUN-1999) Genoscope - Centre National de Sequencage :  
BP 191 91006 EVRY cedex - FRANCE (E-mail : seqrefgenoscope.cns.fr  
- Web : www.genoscope.cns.fr)  
Determination of this BAC-end sequence was carried out as part of a  
collaboration with the Berkeley Drosophila Genome Project (BDGP).  
The BDGP is constructing a physical map of the Drosophila  
melanogaster genome using these BACs. For further information  
please see <http://www.fruitfly.org> The BDGP Drosophila  
melanogaster BAC library was prepared by Kazuhiro Ohsawa and  
Aaron Mammossier in Piter de Jong's laboratory in the Department of  
Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,  
NY. The library is named RPCI-98 and was constructed by partial









[illegible]

FEATURES	source
RESULT 10	
CNS0100X	
LOCUS	
DEFINITION	CNS0100X 1101 bp DNA linear GSS 26-JUL-1999
ACCESSION	Drosophila melanogaster genome survey sequence Sfe end of BAC
VERSION	BACN03G04 of DrosBAC library from Drosophila melanogaster (fruit
KEYWORDS	fly), genomic survey sequence.
SCOURCE	AL098379
ORGANISM	AL098379.1 GI:5609990
REFERENCE	GSS.
AUTHORS	Drosophila melanogaster.
TITLE	Drosophila melanogaster.
JOURNAL	Eukaryote; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
COMMENT	Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
	Ephydroidea; Drosophilidae; Drosophila.
	1 (bases 1 to 1101)
	Genoscope.
	Direct Submission
	Submitted (23-JUL-1999) Genoscope - Centre National de Sequencage; fr
	Bp 191 91006 EVRY cedex - FRANCE (E-mail : seque@genoscope.cns.fr
	- Web : www.genoscope.cns.fr)
	Determination of this BAC-end sequence was carried out as part of a
	collaboration with the European Drosophila Genome Project (EDGP) -
	http://www.edgp.ebi.ac.uk -. This Drosophila melanogaster BAC
	library (Dros BAC) was made by Alain Billaud at CEPH (Centre
	d'Etude du Polymorphisme Humain) with funding provided by a MRC
	project grant. The DNA was prepared from embryos by Alain Bucheton
	and Genevieve Payan. It has been constructed in the vector
	pBelobAC11.
	location/Qualifiers
	1..1101
	/organism="Drosophila melanogaster"
	/db_xref="taxon:7227"
	/clone="BACN03G04"
	/clone_1id="DrosBAC"

```

BASE COUNT      195 a      108 c      131 g      161 t      506 others
ORIGIN
Query Match:      4.2%;      Score 62.4;      DB 17;      Length 1101;
Best Local Similarity 17.4%;      Pred. No. 0.0026;
Matches 107;      Conservative 262;      Mismatches 243;      Indels 2;      Gaps 2

```



Oy	682	GTTATCTGTGTCGTGCATCCTGGTGGTGCTCGTCTACCTAGCCGGTGT	741
Db	455	DTWKTWTWTTTTTGWKKMDTKTKMKDDBSTWTDWMKDKDDTKTWTDWA	514
Oy	742	GTAATTGTCAGACCCTTGATAAATCAACTGCTGGTGGATGACTTTGTGACATTCAGTA	801
Db	515	AGAWMTDKDTAKAKKRTKTDTKTKTRTGTGKKTIDGMRKPTDWGMDRABRTDARAD	574
Oy	802	GCAAGACTATTTCCATCTCGGTCAATTGTGTGGTGAGGTGACGTGATGGAGTG	861
Db	575	RAANAKDXADKADADT-DATKTDGWTTKADTTDDTDADMWDKMWDKADDTKAGR	633
Oy	862	AATTGAGGTAATGGATGGAGTAGAAATGTGATGTGATGGATGATTTGAGACCATTTGA	921
Db	634	GMWGKGKGGWKTKTKTKKKTKTKTGTDPMKWGMDKMTDCKRDGRGGDRDKGTGD	693
Oy	922	GGAGAGAAGACTCATGAATATCTATGCAACGATGGAAGTGGTGTGATCAATGAAGTA	981
Db	694	GBSKAKGDAAMRAKHTPBAADAATAAATAKATMAKTITTKTTTTTTT	753
Oy	982	GGGGTCAAATGTTTGAATGTATGTGTTTGGAGAGTGGTAGATGAG-AGTGACTTGATC	1040
Db	754	TTTTCTTTTARRADRDDDKAKRDEGKDMDADKDMFARGDRBDPTRKMGKORAKX	813
Oy	1041	GTCCGAATATAGGCACATGTGATGTGATGTGACATATTGTTGTGTGGTAGTGAAGG	1100
Db	814	KAKKKDHHTAAMWADDTGTAWAYAKVTATATGKAKAKBKRGGGGGGRDGRAGRR	873
Oy	1101	TGGATATTTGTCGACGTTGATATTTTGAATATCACCTGTGTGTTTGTGTCGGTG	1160
Db	874	KRARGRKRRRKDKRKRKTKKAGDGTGAGTDPAKDGKKAADGMMKMKMSKTK	933
Oy	1161	TGTTTGTCTTTCTCATCTGTACTGTACGTTGTTGTACTGTACTGTTGTGTTGT	1220
Db	934	TMTGTDPTTTWKGNRRGRGAGARRGRGKAGKAGGMMWAMVADAWMTWMDK	993
Oy	1221	TGTTGCCACACAGGTCGACGACATATATGATGATGCCGACATGCGAACATGCGATC	1280
Db	994	DKRWKTMWGMGKGGWGDGRKGGADGAKRNKNRDNRYGNRMWMDAATAVARWAD	1053
Oy	1281	CTATATGTGTGAAA	1294
Db	1054	DDMADYDDDDAD	1067
RESULT 11			
CNS017ZT			
LOCUS		1101 bp DNA linear GSS 26-JUL-1999	
DEFINITION		Drosophila melanogaster genome survey sequence T7 end of BAC	
		BACN3JN10 of DrosBAC library from Drosophila melanogaster (fruit	
		fly), genomic survey sequence.	
ACCESSION		AL108707	
VERSION		ALI08707.1 GI:5629011	
KEYWORDS		GSS.	
SOURCE		Drosophila melanogaster.	
ORGANISM		Drosophila melanogaster	
		Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;	
		Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;	
		Ephydroidea; Drosophilidae; Drosophila.	
REFERENCE		1 (bases 1 to 1101)	
AUTHORS		Genoscope.	
TITLE		Direct Submision	
JOURNAL		Submitted (23-JUL-1999) Genoscope - Centre National de Sequenage :	
		BP 191 91006 EVRY cedex - FRANCE (E-mail : sequefgenosco.cns.fr	
COMMENT		- Web : www.genoscope.cns.fr)	
		Determination of this BAC-end sequence was carried out as part of a	
		collaboration with the European Drosophila Genome Project (EDGP) -	
		http://www.edgp.ebi.ac.uk -. This Drosophila melanogaster BAC	
		library (Dros BAC) was made by Alain Billaud at CBPH (Centre	
		d'Etude du Polymorphisme Humain) with funding provided by a MRC	
		project grant. The DNA was prepared from embryos by Alain Bucheton	
		and Genevieve Payan. It has been constructed in the vector	
		pBelobAC11.	

FEATURES	location/Qualifiers
source	1..1101
	/organism="Drosophila melanogaster"
	/db_xref="taxon:7227"
	/clone="BACN37N10"
	/clone_1ib="DrosBAC"
	/plasmid="pBelOBAC11"
	/note="end : 17"
BASE COUNT	193 a 136 c 251 g 245 t 276 others
ORIGIN	
Query Match	4.2% Score 62.4; DB 17; Length 1101;
Best Local Similarity	27.8%; Pred. No. 0.0026; Mismatches 183; Indels 0; Gaps 0;
Matches 111; Conservative 105; Mismatches 183; Indels 0; Gaps 0;	
827	TGTGTGGAGAGTAGGAGCAGTATGTGAGGTGAGTTGAGTAGTATGATGATGAGATGG 886
Db	657 TGAATTTGGKKAAGTGTGKATATPAKAKTTTMDWAMGAMRRMRDRTRGGGKGTATR 716
Qy	887 AATGTATGTATGAGTGGATGATTGAGACCACTTGAGAGAGAGAAGACATGATAATATCTA 946
Db	717 KQAGARRKRRARRRRMTTGTGRRGTAATRTTKTGKRRPWTNAAAAA 776
Qy	947 TGCACCAATGAGATGATGTTGTGTATCATGAAAGTAGGGGTCATATGTTTGAATATGTG 1006
Db	777 GBDGATGKTRKAKKTKTKGKGGGGGKKTGTTKGTGGGGGTGDMGWTGKTKTKTTT 836
Qy	1007 TTTGAGAGTGGTGGAGAGAGAGTATGATCGTGAATATAGTGCACATGTGATTTGT 1066
Db	837 KTTTTTGTTTTGTGTGTGGGSDGTTTKKTTTTKKGGTGTGKGGGGG 896
Qy	1067 ATGTGACTATTTGTGTGTGGTGTGTAAGAGGTGATATTTGTCAGTTGATATTTTC 1126
Db	897 GGTGTGTGTGTTGTGKGTGTGKGTGTGKGTGKGTGKGTGKGTGKGTGKGTGKGTG 956
Qy	1127 GAAATTTCACTGTGTGTTTGTGTTTCCGTGTGTTTGTGTTTCTTTCATCTGACTG 1186
Db	957 KKKKGGGTGKGGKGTGKGTGKGTGKGTGKGTGKGTGKGTGKGTGKGTGKGTGKGTG 1016
Qy	1187 TACGTGTGTGATCTGACTGTTGTGTGAGTTGTTGTTT 1225
Db	1017 KGTGTGKGTGTTGTTGKGTGKGTGKGTGKGTGKGTGKGTGKGTGKGTGKGTGKGTG 1055
RESULT 12	
CNS0166K	
LOCUS	CNS0166K 1225 bp DNA linear GSS 26-JUL-1999
DEFINITION	Drosophila melanogaster genome survey sequence spe end of BAC BACN37N10 of DrosBAC library from Drosophila melanogaster (fruit fly), genomic survey sequence.
ACCESSION	AL106358
VERSION	AL106358.1 GI:5621512
KEYWORDS	GSS.
SOURCE	Drosophila melanogaster.
ORGANISM	Drosophila melanogaster
	Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE	1 (bases 1 to 1225)
AUTHORS	Genoscope.
TITLE	Direct Submission
JOURNAL	Submitted (23-JUL-1999) Genoscope - Centre National de Sequencage : BP 191 91006 EVRY cedex - FRANCE (E-mail : seqret@genoscope.cns.fr - Web : www.genoscope.cns.fr)
COMMENT	Determination of this BAC-end sequence was carried out as part of a collaboration with the European Drosophila Genome Project (EDGP) - http://www.edgp.ebi.ac.uk -. This Drosophila melanogaster BAC library (Dros BAC) was made by Alain Billaut at CEPH (Centre d'Etude du Polymorphisme Humain) with funding provided by a MRC project grant. The DNA was prepared from embryos by Alain Bucheton and Genevieve Payan. It has been constructed in the vector pBelOBAC11.
FEATURES	location/Qualifiers





Db 699 DGADAGKGGKTKRRRRDRATWDRDAMWADAMWTTTDTDDWDKDRRRKCARRRR 640

Qy 964 TGTGTGATCATGAAGTAGGGGTCAATGTG-----TTGAGATAGTGTGGAGAGTGG 1018

Db 639 RTTAAAMAMWMTWKAMDWKMCKTRADKWDKADTWDADKADBDKAKAMRRRRR 580

Qy 1019 TGAATGAGAGTGAAGTTCATGTCGATATAGTGCACATGATGATGATGAGACTATT 1078

Db 579 AAAAARBRWTTTKKTTTATTTWTTMABAAAMWAMWATTTATTTTWTWTWTWT 520

Qy 1079 GTTGTGTGGGTGTGTAAAGGCTGATTCGCCAGTTGATTTTGCAGAAATTCCTTG 1138

Db 519 TTTWTTTAAAMWMTATATAAWTAAAMAAAAAAATTTTTTTTTTTTTTWTAAWMTA 460

Qy 1139 TGTGTTTGTGTTTGTCCGTTGTTGTTGTTCTTCATCTGTA 1183

Db 459 WTTTWTWTWTWTAAATTTTWTWTWTAAATTTTWTWTWT 415

RESULT 16  
A0743348/c 1114 bp DNA linear GSS 16-JUL-1999

LOCUS HS\_5387\_B2.F11\_SPE\_RPCL-11 Human Male BAC Library Homo sapiens  
genomic clone Plate=963 Col=22 Row=L, DNA sequence.

ACCESSION A0743348  
VERSION A0743348.1 GI:5520870  
KEYWORDS GSS.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
1 (bases 1 to 1114)  
Mahiwas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T.,  
Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and  
Hood,L.  
Sequence-tagged connectors: A sequence approach to mapping and  
scanning the human genome  
Proc. Natl. Acad. Sci. U.S.A. 96 (17), 9739-9744 (1999)  
99380589

TITLE Contact: Mahiwas G.G., Wallace J.C., Hood L  
JOURNAL High Throughput Sequencing Center  
MEDLINE University of Washington  
COMMENT 401 Queen Anne Avenue North, Seattle, WA 98109, USA  
Tel: (206) 616-3618  
Fax: (206) 616-3887  
Email: jwallace@u.washington.edu  
Clones are derived from the human BAC library RPCL-11. For BAC  
library availability, please contact Pieter de Jong  
(pieter@dejong.med.bufileo.edu). Clones may be purchased from  
BACPAC Resources ([http://bacpac.med.bufileo.edu/ordering\\_bac.htm](http://bacpac.med.bufileo.edu/ordering_bac.htm))  
or from Research Genetics ([info@resgen.com](http://info@resgen.com)). BAC end Web Server:  
<http://www.htsc.washington.edu>  
Plate: 963 row: L column: 22  
Seq primer: SPE  
Class: BAC ends  
High quality sequence stop: 1114.  
Location/Qualifiers  
1..1114  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="Plate=963 Col=22 Row=L"  
/clone\_lib="RPCL-11 Human Male BAC Library"  
/sex="male"  
/note="Vector: pBAC3.6, Site 1: EcoRI, Site 2: EcoRI;  
Male blood DNA was isolated from one randomly chosen donor  
and partially digested with a combination of EcoRI and  
EcoRI Methylase. Size selected DNA was cloned into the  
pBAC3.6 vector at EcoRI sites"

BASE COUNT 634 a 393 c 44 g 6 t 37 others

ORIGIN

Query Match 4.0%; Score 59.2; DB 17; Length 1114;  
Best Local Similarity 45.4%; Pred. No. 0.013;

Matches 240; Conservative 0; Mismatches 288; Indels 1; Gaps 1;

Qy 686 TTTCTGTGTCTGTATCTATCTTTGTGTGTCTGTCTACTACCTGATCCGTTTGTAT 745

Db 831 TTTTGTGTTCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 772

Qy 746 TGTTCAGAGCCTGTAAATTAACAACCTGTGTGTGTGTGTGTGTGTGTGTGTGT 805

Db 771 TGTTTTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 712

Qy 806 AGTATTTTCCATCTGCGCATCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 865

Db 711 TGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 652

Qy 866 GAGGTAGATGTGATGAGATGAGATGAGATGAGATGAGATGAGATGAGATGAGAT 925

Db 651 GTTTTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 592

Qy 926 AGAAGATCTGAAATATCTATGCAAGATGAAAGTGTGTGTGTGTGTGTGTGT 985

Db 591 TGTTTTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 532

Qy 986 GTCAATGCTTTGATGATGATGATGATGATGATGATGATGATGATGATGAT 1045

Db 531 GTTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 473

Qy 1046 ATATAGTGACATGATGATGATGATGATGATGATGATGATGATGATGAT 1105

Db 472 TGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 413

Qy 1106 ATGTGTCAGTTGATATTTTGCAGAAATCACTGTGTGTGTGTGTGTGTGT 1165

Db 412 GTTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 353

Qy 1166 TGTTCCTTCTCATCTGACTGACCTGTGTGTGTGTGTGTGTGTGTGTGT 1214

Db 352 GGTTTTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 304

RESULT 17  
CNS01844/c 1101 bp DNA linear GSS 26-JUL-1999

LOCUS CNS01844  
DEFINITION Drosophila melanogaster genome survey sequence T7 end of BAC  
BACN37D10 of DrosBAC library from Drosophila melanogaster (fruit  
fly), genomic survey sequence.  
AL108862

ACCESSION AL108862.1 GI:5629166  
VERSION  
KEYWORDS  
SOURCE Drosophila melanogaster.  
ORGANISM Drosophila melanogaster  
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;  
Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
Ephydroidea; Drosophilidae; Drosophila.  
1 (bases 1 to 1101)

REFERENCE  
1 Direct Submission  
Genoscope.  
Submitted (23-JUL-1999) Genoscope - Centre National de Sequencage :  
BP 191 91006 EVRY cedex - FRANCE (E-mail : [seql@genoscope.cns.fr](mailto:seql@genoscope.cns.fr)  
- Web : [www.genoscope.cns.fr](http://www.genoscope.cns.fr))  
determination of this BAC-end sequence was carried out as part of a  
collaboration with the European Drosophila Genome Project (EDGP) -  
<http://www.edgp.ebi.ac.uk> - This Drosophila melanogaster BAC  
library (Dros BAC) was made by Alain Billand at CESP (Centre  
d'Etude du Polymorphisme Humain) with funding provided by a MRC  
project grant. The DNA was prepared from embryos by Alain Bucheton  
and Genevieve Payan. It has been constructed in the vector  
pBelBAC11.

FEATURES  
source 1..1101  
Location/Qualifiers  
/organism="Drosophila melanogaster"  
/db\_xref="taxon:7227"  
/clone="BACN37D10"  
/clone\_lib="DrosBAC"





REFERENCE	AUTHORS	TITLE	JOURNAL
Ephydroidae; Drosophilidae; Drosophila.	1 (bases 1 to 1101)	Genoscope.	
Direct Submission			
Submitted (23-JUL-1999) Genoscope - Centre National de Sequencage :			
BP 191 91006 Evry cedex - FRANCE (E-mail : seqref@genoscope.cns.fr			
- Web : www.genoscope.cns.fr)			
Determination of this BAC-end sequence was carried out as part of a			
collaboration with the European Drosophila Genome Project (EDGP) -			
http://www.edgp.ebi.ac.uk -. This Drosophila melanogaster BAC			
library (Dros BAC) was made by Alain Billard at CRPH (Centre			
d'etude du Polymorphisme Humain) with funding provided by a MRC			
project grant. The DNA was prepared from embryos by Alain Bucheton			
and Genevieve Payan. It has been constructed in the vector			
pBelobAC11.			
FEATURES	source	Location/Qualifiers	
	1..1101	/organism="Drosophila melanogaster"	
		/db_xref="taxon:7227"	
		/clone="BACN03K20"	
		/clone_11b="DrosBAC"	
		/plasmid="pBelobAC11"	
		/note="end : 17"	
BASE COUNT	258 a	107 c	60 g 175 t 501 others
ORIGIN			
Query Match	4.0%; Score 58.4; DB 17;	Length 1101;	
Best Local Similarity	18.8%; Pred. No. 0.019;		
Matches 90; Conservativity 209; Mismatches 179; Indels 2; Gaps 2;			
Gy	654	GACTGAGTCGTCACTGATGCATGTCACATGTTCTGTGTCGTGTGA-TCTGTTTGTG	712
Db	539	KMCKCTGTGGTGTCASAGRTMGKVXBKHHTHMYTGTKHCTCHHKITTKTWTTTTM	598
Gy	713	TGTCCTGTGTCACCTGACCGTGTGTTGTATTGTGTGTCGAGCCTTGATAAACAACG	772
Db	599	TMKAGTGTGTTTDAKATRAITAATGKTATRTTAATATKAARADPKDRPTAATAAKATK	658
Gy	773	TGTTTGATGACTTTGTGTGACAGTTGACAGTAGAAGTTCATTCGTCGTCATTGTGT	832
Db	659	KKARKDKTAPADAMWMDRATRKAKKKAAKAKAKAAATATGDKARAKAKAKATBA	718
Gy	833	GGTAGAGTGAGGTGACCGTGATGTGAGTGAAGTGAAGTGAATGCAATGTG	892
Db	719	KARAWDTATWTATADAADAKGAKAKAKAVADRKRMRPAKDKRRARAKAKAAD	778
Gy	893	ATGTGATGGATGATTTGAGAACAATTGGAGAGAGAGAACGATGATAATATCATGAAA	952
Db	779	DADADAKAKAADAAKADADADJRGSGKKRKAADKKKKKKAKAGDDKKAKAKDRA	838
Gy	953	CGATGGAAGTGTGTGTATCAT-GAAGTAGAGGGGTCATGTGTTAGTATGTGTTGG	1011
Db	839	AKAKADAAGDAKAKARRRAGDDKAKADAKAKAKAKADDKATTKAKATKAKGDAX	898
Gy	1012	AGATGTGAGATGAGAGAGTGACTTGATGTCGAAATATATGTCACATGATGTATGTG	1071
Db	899	AKKKKKDKAKAKAKAKAKADKDDKDDKDDKADAKKKKKDKDRAKAKKDDDK	958
Gy	1072	GACTATTTGTTGTGTCGGATGTTAAAGGTGATATTGTGCCAGTGTATTTTGAAT	1131
Db	959	DKAKDKDKDKADAKAAKADADAKAKAKAADDDAADAADAKADAKAKAKAK	1018
RESULT 22	LOCUS	DEFINITION	VERSION
AQ743361/c	AQ743361	HS_5387_B2_H05_SP6_RPCI-11 Human Male BAC Library Homo sapiens genomic clone Plate=963 Col=10 Row=P, DNA sequence.	AQ743361
ACCSSION	AQ743361		
VERSION	AQ743361.1	GI:5520883	
KEYWORDS	GSS.		
SOURCE	human.		
ORGANISM	Homo sapiens		

REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
AUTHORS	1 (bases 1 to 1069) Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T., Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and Hood,L.
TITLE	Sequence-tagged connectors: A sequence approach to mapping and scanning the human genome
JOURNAL	Proc. Natl. Acad. Sci. U.S.A. 96 (17) , 9739-9744 (1999)
MEDLINE	99380589
COMMENT	Contact: Mahairas GG, Wallace JC, Hood L High Throughput Sequencing Center University of Washington 401 Queen Anne Avenue North, Seattle, WA 98109, USA Tel.: (206) 616-3618 Fax: (206) 616-3887 Email: jwallace@u.washington.edu Clones are derived from the human BAC library RPCI-11. For BAC library availability, please contact Pieter de Jong (piet@dejong.med.buffalo.edu). Clones may be purchased from BACPAC Resources ( <a href="http://bacpac.med.buffalo.edu/ordering_bac.htm">http://bacpac.med.buffalo.edu/ordering_bac.htm</a> ) or from Resear h Genetics ( <a href="http://inforesgen.com">http://inforesgen.com</a> ). BAC end Web Server: <a href="http://www.htbc.washington.edu">http://www.htbc.washington.edu</a> Plate: 963 row: P column: 10 Seq primer: SP6 Class: BAC ends High quality sequence stop: 1069. Location/Qualifiers 1..1069 /organism="Homo sapiens" /db_xref="taxon:9606" /cclone="Plate=963 Col=10 Row=P" /clone_lib="RPCI-11 Human Male BAC Library" /sex="male" /note="Vector: pBACe3.6; Site 1: EcoRI; Site 2: EcoRI; Male blood DNA was isolated from one randomly chosen donor and partially digested with a combination of EcoRI and EcoRI Methylase. Size selected DNA was cloned into the pBACe3.6 vector at EcoRI sites"
BASE COUNT	696 a 275 c 42 g 10 t 46 others
ORIGIN	
Query Match	3.8%; Score 56.4; DB 17; Length 1069;
Best Local Similarity	52.3%; Pred. No. 0.053;
Matches 123; Conservative	0; Mismatches 112; Indels 0; Gaps 0;
Oy	992 GTGTTTGAGTATGTGGTAGAGTGGAAGATGAGACTGTATCGCAATTAG 1051
Db	657 GTCGTGTTTTTTTTTTTTTGTGGTTTTGTGTTGTGTGTGTGTGTTTGTG 598
Oy	1052 TGCACATGTGATTTGATGTGACATACTATTGTTGTGTGGTAGTGAAGGATGATATGTG 1111
Db	597 TGGGGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 538
Oy	1112 CCAATTGATTTTTCGAATTCACCTGTGTGTGTGTGTGTGTGTGTGTGTGTCT 1171
Db	537 TGTGTGTGCCTTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 478
Oy	1172 TTTCACATGTGACGTGACGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTG 1226
Db	477 TTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 423

RESULT 23

CNS000D1                  1101 bp     DNA       linear     GSS 03-JUN-1999

DEFINITION              Drosophila melanogaster genome survey sequence TEI3 end of BAC #

                          BACR01J16 of RPCI-98 library from Drosophila melanogaster (fruit fly), genomic survey sequence.

ACCESSION                AL065414

VERSION                  AL065414.1    GI:4938827

KEYWORDS                GSS.

SOURCE                  Drosophila melanogaster.



[illegible]

LOCUS	BM415882	967 bp	RNA	linear	EST 28-JAN-2002
DEFINITION	Op20964 Mixed Stage EST's from Globodera pallida, the potato cyst nematode Globodera pallida cDNA, mRNA sequence.				
ACCESSION	BM415882				
VERSION	BM415882.1	GI:18382681			
KEYWORDS	EST.				
SOURCE	Globodera pallida.				
ORGANISM	Eukaryota; Metazoa; Nematoda; Chromadorea; Tylenchida; Tylenchina; Tylenchoidea; Heteroderidae; Heteroderinae; Globodera.				
REFERENCE	1 (bases 1 to 967)				
AUTHORS	Heer J., Sosinski B., Pokrzywa R.M., Warry A. and Opperman C.				
TITLE	Mixed Stage EST's from Globodera pallida, the potato cyst nematode				
JOURNAL	Unpublished (2001)				
COMMENT	Contact: Opperman, C Center for the Biology of Nematode Parasitism NC State University; IACR-Rothamsted Campus Box 7616; Raleigh, NC 27695, USA Tel: 919.515.6689 Fax: 919.515.9500 Email: warthog@unity.ncsu.edu G111-8PCN R A05 G111-8 R 033.ab1. Location/Qualifiers				
FEATURES	1. .967				
SOURCE	/organism="Globodera pallida" /db_xref="taxon:36090" /clone_id="Mixed Stage EST's from Globodera pallida, the potato cyst nematode" /note="Vector: lambda G111. This is a collaborative effort between IACR-Rothamsted and North Carolina State University. The library was constructed from mixed stage G. pallida in lambda G111 by Paul Burroughs, IACR-Rothamsted."				
BASE COUNT	500 a 190 c 14 g 251 t 12 others				
ORIGIN					
Query Match	3.8%; Score 56; DB 13; Length 967;				
Best Local Similarity	50.8%; Pred. No. 0.064;				
Matches 134; Conservative	0; Mismatches 130; Indels 0; Gaps 0;				
Qy	962 TGTGTTGGTACATGAAGAGGGGGCTCAATGCTGTGAGTATGTTTGGAGAGGTGA 1021				
Db	923 TGTGGGTGTATGTTTTTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 864				
Qy	1022 GATGAGAGTGACTTGATCGTGCATATATAGTCACATSGATGTATGATGACATATGTT 1081				
Db	863 GTTGGGTGTGTATTTTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 804				
Qy	1082 GTGTGGGTAGTGAAGGTGATATGTGCCAGTTATTTTGGAAATTCACTTGCT 1141				
Db	803 AATGATATTTTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 744				
Qy	1142 GTTTTGTTTGTTCGTTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1201				
Db	743 TTTTGTTTTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 684				
Qy	1202 TACTGTTGTGTGTGTGTGTGT 1225				
Db	683 TTTTGTTTTGTGTGTGTGTGTGT 660				
RESULT 25					
CNS003OT/c					
LOCUS	CNS003OT	948 bp	DNA	linear	GSS 03-JUN-1999
DEFINITION	Drosophila melanogaster genome survey sequence T7 end of BAC # BAC08b1 of RpCl-98 library from Drosophila melanogaster (fruit fly), genomic survey sequence.				
ACCESSION	AL064777				
VERSION	AL064777.1	GI:4942331			
KEYWORDS	GSS.				
SOURCE	Drosophila melanogaster.				
ORGANISM	Drosophila melanogaster				







[illegible]

LOCUS	CNS0171b	987 bp	DNA	linear	GSS 26-JUL-1999
DEFINITION	Drosophila melanogaster genome survey sequence Sp6 end of BAC BACN37A07 of DrosBAC library from Drosophila melanogaster (fruit fly), genomic survey sequence.				
ACCESSION	AL108185				
VERSION	AL108185.1 GI:5628489				
KEYWORDS	GSS.				
SOURCE	Drosophila melanogaster.				
ORGANISM	Drosophila melanogaster. Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.				
REFERENCE	1 (bases 1 to 987) Genoscope. Direct Submission Submitted (23-JUL-1999) Genoscope - Centre National de Sequencage : BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr - Web : www.genoscope.cns.fr) Determination of this BAC-end sequence was carried out as part of a collaboration with the European Drosophila Genome Project (EDGP) - http://www.edgp.ebi.ac.uk -. This Drosophila melanogaster BAC library (Dros BAC) was made by Alain Billaud at CEPH (Centre d'Etude du Polymorphisme Humain) with funding provided by a MRC project grant. The DNA was prepared from embryos by Alain Bucheton and Genevieve Payan. It has been constructed in the vector pBelosBAC11.				
FEATURES	Location/Qualifiers 1..987 /organism="Drosophila melanogaster" /db_xref="taxon:7227" /clone="BACN37A07" /clone_1fb="DrosBAC" /plasmid="pBelosBAC11" /note="end : SP6"				
BASE COUNT	224 a	122 c	180 g	211 t	250 others
ORIGIN					
Query Match	3.7%; Score 55; DB 17; Length 987;				
Best Local Similarity	32.0%; Pred. No. 0.11;				
Matches 116; Conservative	87; Mismatches 153; Indels 6; Gaps 1;				
QY	863	GTTCAGGTGACATGGATGGCATGAATGTGATGTGATGGATGGATGGATGACACCTGGAG	922		
Db	632	GGTGAGTCTTCKKKGKGGTWTTWAMWKGKKGKGGGDKDGKWTGYTTCVAGGA	691		
QY	923	GAGAGAGACTCATGAATATCTATGCAACAGATGGAAGTGTGTATCATGAAATAG	982		
Db	632	GGACGAAACGADSGKMWAAANGRAAAAATTTTMMAMTHNHTARDAWMAAAA	751		
QY	983	GGGGTCAATGTGTTGATATGTGTTGGAGAGTGGTGAATGAGAGTGAATGATCGT	1042		
Db	752	WGR-----GTGTAKAKATDTBYMTWGGGGGKKDGGTGGKMKMAWMDHMDTKATAAW	805		
QY	1043	CGAATATGTCACATGTGATGTATGTGACATATGTTGTGTGGGTAGCTAAAGGTG	1102		
Db	806	KAAATATAAAAAAMATKTMTMTAAWTKMTKATKTGTTTAAWMDATGCGKGGCGSD	865		
QY	1103	GATATGTGACAGATGATATTTTGAATTCACATCTGTGTGTTGTTGTTGTTGTCGTTGTG	1162		
Db	866	RAAAAAAGMRAMATTKDTTBTTTWTMTWYTTTCHTTTTTTTTTTTTTTTTTTTT	925		
QY	1163	TTTTGTTCTTCATCTGATGATGATGATGATGATGATGATGATGATGATGATGATG	1222		
Db	926	TYTATTKYTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT	985		
QY	1223	TT 1224			
Db	986	YB 987			
RESULT 31					
CNS009KS/c	791 bp DNA linear GSS 03-JUN-1999				
LOCUS					

DEFINITION Drosophila melanogaster genome survey sequence T7 end of BAC # BACR19F04 of RPCI-98 library from Drosophila melanogaster (fruit fly), genomic survey sequence.

ACCESSION AL053801.1 GI:4935176

VERSION GSS.

KEYWORDS Drosophila melanogaster.

ORGANISM Drosophila melanogaster.

SOURCE Drosophila melanogaster.

REFERENCE Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.

AUTHORS 1 (bases 1 to 791)

TITLE Genoscope.

JOURNAL Direct Submission

COMMENT Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage : BP 191 91006 EVRY cedex - FRANCE (E-mail : seqrefgenoscope.cns.fr - Web : www.genoscope.cns.fr)

Determination of this BAC-end sequence was carried out as part of a collaboration with the Berkeley Drosophila Genome Project (BDGP). The BDGP is constructing a physical map of the Drosophila melanogaster genome using these BACs. For further information please see <http://www.fruitfly.org/TheBDGP/Drosophila>

melanogaster BAC library was prepared by Kazuo Oosawa and Aaron Mammosser in Pieter de Jong's laboratory in the Department of Cancer Genetics at the Roswell Park Cancer Institute in Buffalo, NY. The library is named RPCI-98 and was constructed by partial EcoRI digestion of Drosophila DNA provided by the BDGP from the isogenic strain Y2; cn bw sp, the same strain used for the BDGP's P1 and EST libraries. A more detailed description of the library and how to order individual BAC clones, the entire library, or filters for hybridization from the BACPAC Resource Center can be found at [http://bacpac.med.buffalo.edu/drosophila\\_bac.htm](http://bacpac.med.buffalo.edu/drosophila_bac.htm).

Location/Qualifiers

1..791

/organism="Drosophila melanogaster"

/db\_xref="taxon:7227"

/clone="BACR19F04"

/clone\_1ib="RPCI-98"

/note="end : 77"

BASE COUNT 464 a 105 c 67 g 78 t 77 others

ORIGIN

Query Match 3.7%; Score 54.8; DB 17; Length 791;

Best Local Similarity 45.7%; Pred. No. 0.11;

Matches 107; Conservative 20; Mismatches 107; Indels 0; Gaps 0;

QY 991 TGTGTTGATGATGTTGGAGAGTGAGATGAGAGTCACTGATCGTGGAATATA 1050

DB 275 KGTGTTGATGATGTTGGAGAGTGAGATGAGAGTCACTGATCGTGGAATATA 216

QY 1051 GTGCACATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1110

DB 215 KGTGTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 156

QY 1111 GCGAGTGAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1170

DB 155 TTTTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 96

QY 1171 TTTTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1224

DB 95 TTTTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 42

RESULT 32

CNS00033/c 1068 bp DNA linear GSS 03-JUN-1999

LOCUS

DEFINITION Drosophila melanogaster genome survey sequence T7 end of BAC: BACR39E22 of RPCI-98 library from Drosophila melanogaster (fruit fly), genomic survey sequence.

ACCESSION AL076666.1 GI:4956244

VERSION GSS.

KEYWORDS Drosophila melanogaster.

SOURCE

ORGANISM Drosophila melanogaster

Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.

REFERENCE 1 (bases 1 to 1068)

AUTHORS Genoscope.

TITLE Direct Submission

JOURNAL Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage : BP 191 91006 EVRY cedex - FRANCE (E-mail : seqrefgenoscope.cns.fr - Web : www.genoscope.cns.fr)

Determination of this BAC-end sequence was carried out as part of a collaboration with the Berkeley Drosophila Genome Project (BDGP). The BDGP is constructing a physical map of the Drosophila melanogaster genome using these BACs. For further information please see <http://www.fruitfly.org/TheBDGP/Drosophila>

melanogaster BAC library was prepared by Kazuo Oosawa and Aaron Mammosser in Pieter de Jong's laboratory in the Department of Cancer Genetics at the Roswell Park Cancer Institute in Buffalo, NY. The library is named RPCI-98 and was constructed by partial EcoRI digestion of Drosophila DNA provided by the BDGP from the isogenic strain Y2; cn bw sp, the same strain used for the BDGP's P1 and EST libraries. A more detailed description of the library and how to order individual BAC clones, the entire library, or filters for hybridization from the BACPAC Resource Center can be found at [http://bacpac.med.buffalo.edu/drosophila\\_bac.htm](http://bacpac.med.buffalo.edu/drosophila_bac.htm).

Location/Qualifiers

1..1068

/organism="Drosophila melanogaster"

/db\_xref="taxon:7227"

/clone="BACR39E22"

/clone\_1ib="RPCI-98"

/note="end : 77"

BASE COUNT 509 a 195 c 5 g 187 t 172 others

ORIGIN

Query Match 3.7%; Score 54.6; DB 17; Length 1068;

Best Local Similarity 35.8%; Pred. No. 0.13;

Matches 221; Conservative 69; Mismatches 326; Indels 1; Gaps 1;

QY 609 TGTGTAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 668

DB 1067 TGTGTAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1008

QY 669 CAGATGTCAGTCACTGATGATGATGATGATGATGATGATGATGATGATGATGAT 728

DB 1007 TGTGTAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 948

QY 729 TGTGTAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 788

DB 947 TGTGTAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 888

QY 789 TGTGTAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 848

DB 887 TGTGTAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 828

QY 849 TGTGTAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 908

DB 827 TGTGTAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 768

QY 909 TGTGTAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 968

DB 767 TGTGTAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 708

QY 969 TGTGTAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1028

DB 707 TGTGTAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 648

QY 1029 TGTGTAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1088

DB 647 TGTGTAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 588

QY 1089 TGTGTAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1148

[illegible]

OY	1007	TTGGAGAGGTGATGAGACAGAGACTTATCGTCGAATATAGTGACAATGATTGT	1066
Dd	921	TTTTTTTTTTGTTTGGKGGGCTTTTTTTTTTKTGCGTTCGTGTCGTGATGTTGGTGGGT	862
OY	1067	ATGTGCACTATTTGTTGTCGTGATGTAAAGGCTGCATATTTCGCCAGTTGATATATTTTC	1126
Dd	861	GTTKMTWTGTTTTTTTKTGTTGTAAGTTGTGKKGGGGGKGKGTGKTGTGGGGGTGKG	802
OY	1127	GAATTCACCTTGTCGTGTTTTGTTTCCGTGTGTTTTGTTCTTCATCTGACTG	1186
Dd	801	GTGTGTGTGTGGKGGTGTGTTGTGTGTGTGTGTGTGTGTGGGGGTGTGATGTGTGTG	742
OY	1187	TACGTTGTGTACGTGACTGTCGTGTTGCGTTGTTGTT	1224
Dd	741	TKTGTGNGGTGTGTGTGTGTGTGTTTGTGTTTGTGK	704
<hr/>			
RESULT 34			
AG135074/c			
LOCUS	AG135074	963 bp	DNA linear GSS 04-NOV-2001
DEFINITION	Pan troglodytes DNA, clone: PTB-148D21.R, genomic survey sequence.		
ACCESSION	AG135074		
VERSION	AG135074.1	GI:16664752	
KEYWORDS	GSS.		
SOURCE	Pan troglodytes male lymphoblast DNA, clone_lib:PTB Chimpanzee Male		
ORGANISM	BAC library, clone:PTB-148D21.R.		
REFERENCE	Pan troglodytes		
AUTHORS	Mumukota; Metaxoa; Chordata; Carniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Pan.		
JOURNAL	1 Fujiyama,A., Hattori,M., Toyoda,A., Taylor,T.D., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.		
REFERENCE	BAC end sequences of library PTB		
AUTHORS	Unpublished		
TITLE	2 (bases 1 to 963)		
JOURNAL	Fujiyama,A., Hattori,M., Toyoda,A., Taylor,T.D., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.		
TITLE	Direct Submission		
JOURNAL	Submitted (02-AUG-2001) Asao Fujiyama, The Institute of Physical and Chemical Research (RIKEN), Genomic Sciences Center (GSC); 1-7-22 Suehiro-chou, Tsurumi-Ku, Yokohama, Kanagawa 230-0045, Japan (E-mail:chimpesgsc.riken.go.jp, URL:http://hgp.gsc.riken.go.jp/, Tel:81-45-503-9111, Fax:81-45-503-9170)		
COMMENT	Clones are derived from the chimpanzee BAC library PTB This BAC end was generated during the R&d process and may have higher chance of clone tracking errors.		
PRIMERS			
Sequencing:	M13Rev		
LIBRARY			
Vector	: pKS145		
R.Site 1	: SacI		
R.Site 2	: SacI		
Location/Qualifiers			
1..963			
/organism="Pan troglodytes"			
/db_xref="taxon:9598"			
/clone="PTB-148D21.R"			
/sex="male"			
/cell_type="lymphoblast"			
/clone_lib="PTB Chimpanzee Male BAC library"			
BASE COUNT	398 a 393 c 57 g 48 t 77 others		
ORIGIN			
Query Match	3.7%; Score 54.2; DB 17; Length 963;		
Best Local Similarity	47.0%; Pred. No. 0.16;		
Matches 149; Conservative	0; Mismatches 168; Indels 0; Gaps 0;		
OY	937	GAATPATTATGCAAGAAGAGAGAGTGTGTGTACATGAGTAGGGGGCATGTGTT	996
Dd	963	GAGATGTAATGCTTCCGANTGNAGATCGTTGATGTGTGATGTCTGTATGATAGANT	904
OY	997	TGAGTATGCTGTTTGAGAGGTGATGAGATGAGAGTGAATGATCGTGAATATAGTCAC	1056

Db 903 GCGATGGGCTGGGGCGAGCCTGTATTTTNTGTCGCGTGTGTCATTTGCTGTGC 844  
Qy 1057 ATGTATTTGTATGTGACATTTGTTGTGGGTAGTGAAGGTGGATATTTGCGCACT 1116  
Db 843 CTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 784  
Qy 1117 TGATATTTTTCGAATTCATCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTCTC 1176  
Db 783 GGGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 724  
Qy 1177 ATCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1236  
Db 723 TGAGGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 664  
Qy 1237 TCGAGCGACTATATATG 1253  
Db 663 TCGGCGGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 647

## RESULT 35

CNS0100P/c

LOCUS CNS0100P 778 bp DNA linear GSS 26-JUL-1999  
DEFINITION Drosophila melanogaster genome survey sequence T7 end of BAC  
BACN04B15 of DrosBAC library from Drosophila melanogaster (fruit  
fly), genomic survey sequence.

ACCESSION AL099451 GI:5611062

VERSION AL099451.1 GI:5611062

KEYWORDS GSS.

SOURCE Drosophila melanogaster.

ORGANISM Drosophila melanogaster.

REFERENCE Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;  
Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
Ephydroidea; Drosophilidae; Drosophila.  
1 (bases 1 to 778)

Genoscope.

Direct Submission

Submitted (23-JUL-1999) Genoscope - Centre National de Sequencage :  
BP 191 91006 EVRY cedex - FRANCE (E-mail : sequef@genoscope.cns.fr  
- Web : www.genoscope.cns.fr)Determination of this BAC-end sequence was carried out as part of a  
collaboration with the European Drosophila Genome Project (EDGP) -  
http://www.edgp.ebi.ac.uk - . This Drosophila melanogaster BAC  
library (Dros BAC) was made by Alain Billand at CEPH (Centre  
d'Etude du Polymorphisme Humain) with funding provided by a MRC  
project grant. The DNA was prepared from embryos by Alain Bucheton  
and Genevieve Payan. It has been constructed in the vector  
pBelobAC11.

FEATURES

Location/Qualifiers

1..778

/organism="Drosophila melanogaster"

/db\_xref="taxon:7227"

/clone="BACN04B15"

/clone\_1lb="DrosBAC"

/plasmid="pBelobAC11"

/note="end : 77"

BASE COUNT 357 a 100 c 20 g 60 t 241 others

ORIGIN

Query Match 3.7%; Score 54; DB 17; Length 778;  
Best Local Similarity 31.6%; Pred. No. 0.17;  
Matches 149; Conservative 57; Mismatches 264; Indels 2; Gaps 1;

Qy 754 GCGTATATATACACTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 813

Db 749 DMMAGAGATKMGCMARKMSCKGTGAAGTCVAGTTTGAAGRAAGRKGGAKRRTGT 690

Qy 814 CCATCTCGGTCATTTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 873

Db 689 TTTATAGADTKAIVRAARSAGATAGAGAGKTRCAGGKGAARRRAGATATTTTGG 630

Qy 874 TTGATGGAGTGAATGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 933

Db 629 GGAAGTAMTSGKRTTGGAAAPATGKARKKAGKAGAGAGCGGAGAGWRKRRKGG 570  
Qy 924 CATGAATATCTATGCAACAGATGAGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 993  
Db 569 AA-ARRRARTTAKARAKTKTRGAKRTRRAGGNNNGGTGTGTGTGTGTGTGTGTGTGT 512  
Qy 994 GTTGAATGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1053  
Db 511 GTTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 452  
Qy 1054 CACATGTATTTGTATGTGACATTTGTTGTGGGTAGTGAAGGTGGATATTTGCGC 1113  
Db 451 MTNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 392  
Qy 1114 AGTGAATATTTTCGAATTCATCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTCTT 1173  
Db 391 TTTTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 332  
Qy 1174 CTCATCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1225  
Db 331 NGTTATNGTNGTNGTNGTNGTNGTNGTNGTNGTNGTNGTNGTNGTNGTNGTNGTNGT 280

## RESULT 36

CNS0126V/c

LOCUS CNS0126V 916 bp DNA linear GSS 26-JUL-1999  
DEFINITION Drosophila melanogaster genome survey sequence SP6 end of BAC  
BACN07D07 of DrosBAC library from Drosophila melanogaster (fruit  
fly), genomic survey sequence.

ACCESSION AL101185 GI:5612796

VERSION AL101185.1 GI:5612796

KEYWORDS GSS.

SOURCE Drosophila melanogaster.

ORGANISM Drosophila melanogaster.

REFERENCE Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;  
Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
Ephydroidea; Drosophilidae; Drosophila.  
1 (bases 1 to 916)

Genoscope.

Direct Submission

Submitted (23-JUL-1999) Genoscope - Centre National de Sequencage :  
BP 191 91006 EVRY cedex - FRANCE (E-mail : sequef@genoscope.cns.fr  
- Web : www.genoscope.cns.fr)Determination of this BAC-end sequence was carried out as part of a  
collaboration with the European Drosophila Genome Project (EDGP) -  
http://www.edgp.ebi.ac.uk - . This Drosophila melanogaster BAC  
library (Dros BAC) was made by Alain Billand at CEPH (Centre  
d'Etude du Polymorphisme Humain) with funding provided by a MRC  
project grant. The DNA was prepared from embryos by Alain Bucheton  
and Genevieve Payan. It has been constructed in the vector  
pBelobAC11.

FEATURES

Location/Qualifiers

1..916

/organism="Drosophila melanogaster"

/db\_xref="taxon:7227"

/clone="BACN07D07"

/clone\_1lb="DrosBAC"

/plasmid="pBelobAC11"

/note="end : SP6"

BASE COUNT 348 a 101 c 116 g 201 t 150 others

ORIGIN

Query Match 3.7%; Score 54; DB 17; Length 916;  
Best Local Similarity 38.1%; Pred. No. 0.17;  
Matches 109; Conservative 31; Mismatches 146; Indels 0; Gaps 0;

Qy 940 ATATCTATCAACAGATGAGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 999

Db 366 ATATTANTTAAMTTTATGATKTTTANKAMTTTAAATAMAAKAKTATTTTWRDTPATA 307

Qy 1000 GTATGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1059

Db 306 GKKKGKGTAKAAATAKTTAGAAWAAKTTTNNNTWTWTCTTTTNTTATTTTAKAAATT 247



Db	727	TGGTTTTTTTGTGTTTTTTGTTTGTTGTTGTTGTCAGCTGATGTTGTTTTTTTTTTT	796
Oy	1210	TTTGCGTTGTTGTTGTTG	1226
Db	797	TGGTTTTTTTGTGTTGTTG	813
<hr/>			
RESULT 39			
CNS016MS			
LOCUS			
DEFINITION Drosophila melanogaster genome survey sequence SP6 end of BAC			
BACN16E12 of DrosBAC library from Drosophila melanogaster (fruit			
fly), genomic survey sequence.			
AL106942			
AL106942.1 GI:5624615			
VERSION			
KEYWORDS			
SOURCE			
ORGANISM			
Drosophila melanogaster.			
Drosophila melanogaster			
Eukaryota; Metazoa; Archipoda; Hexapoda; Insecta; Preyigota;			
Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;			
Ephydroidea; Drosophilidae; Drosophila.			
1 (bases 1 to 1101)			
Genoscope.			
Direct Submission			
Submitted (23-JUL-1999) Genoscope - Centre National de Sequencage :			
BP 191 91006 EVRY cedex - FRANCE (E-mail : segre@genoscope.cns.fr			
- Web : www.genoscope.cns.fr)			
Determination of this BAC-end sequence was carried out as part of a			
collaboration with the European Drosophila Genome Project (BDGP) -			
http://www.edgp.ebi.ac.uk -. This Drosophila melanogaster BAC			
library (Dros BAC) was made by Alain Billard at CRPH (Centre			
d'Etude du Polymorphisme Humain) with funding provided by a MRC			
project grant. The DNA was prepared from embryos by Alain Bucheton			
and Genevieve Payan. It has been constructed in the vector			
pBelOBAC11.			
FEATURES			
source			
location/Qualifiers			
1..1101			
/organism="Drosophila melanogaster"			
/db_xref="taxon:7227"			
/clone="BACN16E12"			
/clone_lib="DrosBAC"			
/plasmid="pBelOBAC11"			
/note="end : SP6"			
BASE COUNT 193 a 90 c 142 g 500 t 176 others			
ORIGIN			
Query Match 3.6%; Score 53.2; DB 17; length 1101;			
Best Local Similarity 45.3%; Pred.No.0.26;			
Matches 130; Conservative 14; Mismatches 143; Indels 0; Gaps 0;			
Oy	938	AAATATCTATGCACAAGATGGAAGTGTGTGTGATCATGAAGAAGGAGGATCATGTGTT	997
Db	166	AABMTTGTAAGAGCAAGKAKAAMATCTKANTTTTTTACTTGTCGTATTATTTGGTTG	225
Oy	998	GAGTATGTGTTGGAGAGTGTGAGATGAGAGTGAAGTGAAGTGAATATAGTGACA	1057
Db	226	GTTTTTTTCTKTGTTGATGAGTCGCKTTTATTTTTTATTAGTGTTCATAATGTTTGA	285
Oy	1058	TCTGATGTGATGACATATGTGTGTGGGAGTGAAGTGAAGGATGATATGCGCAGTT	1117
Db	286	TATGAAGGGTTTTTTTTTTTTTTTTTTTTTTTTTTTTTATATATTTTGTGATTTTCT	345
Oy	1118	GATATTTTGCAAATCACTTGTCGTGTTTGTGTTGTTCCGTGTGTTGTTGTTCTTCTCA	1177
Db	346	TMTTTTTTKATTTTATTTTATTTTATTTTCTTGATATATTTTTTTTCTTWTATTTT	405
Oy	1178	TCTGTAAGTGAAGTGTGTAAGTGAAGTGTGTTGTTGTTGTTT	1224
Db	406	TTTTATTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT	452
<hr/>			
RESULT 40			
BG030246/C			

LOCUS	DEFINITION	ACCESSION	VERSION	SOURCE	ORGANISM	REFERENCE	AUTHORS	TITLE	JOURNAL	COMMENT
BG030246	602297544F1 NIH_MGC_87 Homo sapiens cDNA clone IMAGE:4391867 5', mRNA sequence.	BG030246	BG030246.1	GI:12419344	human.	NIH-MGC <a href="http://mgc.nci.nih.gov/">http://mgc.nci.nih.gov/</a> . National Institutes of Health, Mammalian Gene Collection (MGC) Unpublished (1999) Contact: Robert Strausberg, Ph.D. Email: <a href="mailto:cgapbs-remail.nih.gov">cgapbs-remail.nih.gov</a> Tissue Procurement: DCTD/DTF	NIH-MGC	NIH-MGC	NIH-MGC	NIH-MGC
1109 bp	linear	EST 24-JAN-2003								
FEATURES	source	Location/Qualifiers	1..1109							
			/organism="Homo sapiens"							
			/db_xref="taxon:9606"							
			/clone="IMAGE:4391867"							
			/clone_1ib="NIH MGC 87"							
			/tissue_type="mammary adenocarcinoma, cell line"							
			/lab_host="DH10B (phage-resistant)"							
			/note="Organ: breast; Vector: pCMV-SPORTS; Site_1: NotI; Site_2: SalI; Cloned unidirectionally; oligo-dT primed. Average insert size 1.383 kb. Library enriched for full-length clones and constructed by Life Technologies. Note: this is a NIH MGC library."							
BASE COUNT	483 a	280 c	210 g	136 t						
ORIGIN										
Query Match	3.6%	Score 53.2	DB 12	Length 1109						
Best Local Similarity	52.2%	Pred. No. 0.26								
Matches 141	Conservative 0	Mismatches 128	Indels 1	Gaps 1						
OY	958	GAAGTGTGTTGTACATGAAGTATAGGGGTCAATGTGTTAGTATGCTTTGGAGAGTG	1017							
DB	662	GAAAGGTGTGTGCGTAGTTAAAGTAGTGTGTATGTGTATGTGTAGAGTGTGTTATTTGTGGG	603							
OY	1018	GTGAGATGAGAGTACTTGATTCGCGAATATAGTGCACATGTGATTTGATGTGACATAT	1077							
DB	602	GTGTGTGTGTGTTGTTATTTGTTCTTATTTGTTATTTGTTATTTGTTGTTGTTGTTGTTGTT	543							
OY	1078	TGTTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT	1137							
DB	542	TGTTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT	483							
OY	1138	GTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT	1196							
DB	482	TTTTTTTTTTTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT	423							
OY	1197	TACTGTACTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT	1226							
DB	422	TTTTTTTTTGCATTTTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT	393							

Search completed: April 8, 2003, 08:24:23  
Job time : 1837 secs

















962 6 1.3 145 4 US-09-496-632C-18 Sequence 18, Appl  
c 963 6 1.3 147 1 US-08-044-547-1 Sequence 1, Appl  
c 964 6 1.3 151 2 US-08-563-148E-6 Sequence 6, Appl  
c 965 6 1.3 151 4 US-09-518-046-21 Sequence 21, Appl  
c 966 6 1.3 151 4 US-09-194-139-1 Sequence 1, Appl  
967 6 1.3 154 2 US-08-710-330A-11 Sequence 11, Appl  
968 6 1.3 154 4 US-09-261-416-5 Sequence 5, Appl  
969 6 1.3 154 4 US-09-261-416-7 Sequence 7, Appl  
970 6 1.3 154 4 US-08-817-787-18 Sequence 18, Appl  
c 971 6 1.3 155 3 US-09-191-647-8 Sequence 8, Appl  
c 972 6 1.3 155 4 US-09-540-245A-8 Sequence 8, Appl  
c 973 6 1.3 156 4 US-09-540-153-8 Sequence 8, Appl  
974 6 1.3 156 4 US-09-261-416-6 Sequence 6, Appl  
975 6 1.3 158 1 US-08-611-107-4 Sequence 4, Appl  
c 976 6 1.3 158 2 US-08-828-832-3 Sequence 3, Appl  
c 977 6 1.3 158 2 US-08-422-560A-4 Sequence 4, Appl  
978 6 1.3 158 4 US-08-468-793-4 Sequence 4, Appl  
c 979 6 1.3 158 4 US-09-267-177-10 Sequence 10, Appl  
c 980 6 1.3 158 4 US-08-679-493A-24 Sequence 24, Appl  
c 981 6 1.3 158 4 US-09-496-632C-16 Sequence 16, Appl  
c 982 6 1.3 159 2 US-08-828-832-1 Sequence 1, Appl  
c 983 6 1.3 159 4 US-09-496-632C-3 Sequence 3, Appl  
c 984 6 1.3 159 4 US-09-496-632C-15 Sequence 15, Appl  
c 985 6 1.3 160 2 US-08-586-272-3 Sequence 3, Appl  
c 986 6 1.3 160 2 US-08-448-438-8 Sequence 8, Appl  
c 987 6 1.3 160 2 US-08-162-402B-14 Sequence 14, Appl  
988 6 1.3 160 3 US-09-082-969-3 Sequence 3, Appl  
989 6 1.3 161 4 US-09-261-416-8 Sequence 8, Appl  
c 990 6 1.3 162 1 US-08-031-399-5 Sequence 5, Appl  
c 991 6 1.3 162 1 US-08-393-305-2 Sequence 2, Appl  
c 992 6 1.3 162 1 US-08-535-733-2 Sequence 2, Appl  
c 993 6 1.3 162 1 US-08-726-817-2 Sequence 2, Appl  
c 994 6 1.3 162 1 US-08-504-042-5 Sequence 5, Appl  
c 995 6 1.3 162 2 US-08-725-969-2 Sequence 2, Appl  
c 996 6 1.3 162 2 US-08-794-524-2 Sequence 2, Appl  
c 997 6 1.3 162 3 US-08-842-947-6 Sequence 6, Appl  
c 998 6 1.3 162 3 US-08-842-947-8 Sequence 8, Appl  
c 999 6 1.3 162 3 US-08-700-651-13 Sequence 13, Appl  
c1000 6 1.3 162 3 US-09-012-515A-2 Sequence 2, Appl

## ALIGNMENTS

RESULT 1  
US-07-956-848A-25  
; Sequence 25, Application US/07956848A  
; Patent No. 5447914  
; GENERAL INFORMATION:  
; APPLICANT: Travis, James  
; APPLICANT: Shafer, William M.  
; APPLICANT: Bangalore, Neelesh  
; APPLICANT: Pohl, Jan  
; TITLE OF INVENTION: Antimicrobial Peptides  
; NUMBER OF SEQUENCES: 59  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Greenlee and Winner, P.C.  
; STREET: 5370 Manhattan Circle, Suite 201  
; CITY: Boulder  
; STATE: Colorado  
; COUNTRY: USA  
; ZIP: 80303  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/07/956,848A  
; FILING DATE: 02-OCT-1992  
; CLASSIFICATION: 514  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Ferber, Donna M.  
; REGISTRATION NUMBER: 33,878

; REFERENCE/DOCKET NUMBER: 3-908  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (303) 499-8080  
; TELEFAX: (303) 499-8089  
; TELEX: 49617824  
; INFORMATION FOR SEQ ID NO: 25:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 21 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: peptide  
; US-07-956-848A-25  
; Alignment Scores:  
; Pred. No.: 10.5 Length: 21  
; Score: 8.00 Matches: 8  
; Percent Similarity: 100.00% Conservative: 0  
; Best Local Similarity: 100.00% Mismatches: 0  
; Query Match: 1.68% Indels: 0  
; DB: 1 Caps: 0  
US-10-020-441-1 (1-1477) x US-07-956-848A-25 (1-21)  
QY 1328 GGTGACACTGCTGCACCTCTCCTC 1351  
|||||  
Db 3 GYASPERGIYGLYPROLEULEU 10  
RESULT 2  
US-08-471-956-25  
; Sequence 25, Application US/08471956  
; Patent No. 5798336  
; GENERAL INFORMATION:  
; APPLICANT: Travis, James  
; APPLICANT: Shafer, William M.  
; APPLICANT: Bangalore, Neelesh  
; APPLICANT: Pohl, Jan  
; TITLE OF INVENTION: Antimicrobial Peptides  
; NUMBER OF SEQUENCES: 59  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Greenlee and Winner, P.C.  
; STREET: 5370 Manhattan Circle, Suite 201  
; CITY: Boulder  
; STATE: Colorado  
; COUNTRY: USA  
; ZIP: 80303  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/471,956  
; FILING DATE: 06-JUN-1995  
; CLASSIFICATION: 514  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/956,848  
; FILING DATE: 02-OCT-1992  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/541,635  
; FILING DATE: 21-JUN-1990  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Ferber, Donna M.  
; REGISTRATION NUMBER: 33,878  
; REFERENCE/DOCKET NUMBER: 3-908  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (303) 499-8080  
; TELEFAX: (303) 499-8089  
; TELEX: 49617824  
; INFORMATION FOR SEQ ID NO: 25:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 21 amino acids  
; TYPE: amino acid



STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
US-08-471-956-25

## Alignment Scores:

Pred. No.:	10.5	Length:	21
Score:	8.00	Matches:	8
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	1.68*	Indels:	0
DB:	1	Gaps:	0

US-10-020-441-1 (1-1477) x US-08-471-956-25 (1-21)

QY 1328 GGTGACAGTGTGACCTCTC 1351

Db 3 GlyAspSerGlyGlyProLeu 10

## RESULT 3

US-08-278-091-14

; Sequence 14, Application US/08278091  
; Patent No. 5506139

GENERAL INFORMATION:

APPLICANT: LOOSMORE, Sheena M

APPLICANT: YANG, Yan-Ping

APPLICANT: CHONG, Pele

APPLICANT: COMEN, Raymond P.

APPLICANT: KLEIN, Michel H.

TITLE OF INVENTION: Analog of Haemophilus Hinf7 Protein with

NUMBER OF SEQUENCES: 23

REDUCED PROTEASE ACTIVITY

CORRESPONDENCE ADDRESS:

ADDRESSEE: Sim &amp; McBurney

STREET: Suite 701, 330 University Avenue

CITY: Toronto

STATE: Ontario

COUNTRY: Canada

ZIP: M5G 1R7

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/278,091

FILING DATE: 21-JUL-1994

CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:

NAME: Stewart, Michael I

REGISTRATION NUMBER: 24,973

REFERENCE/DOCKET NUMBER: 1038-371

TELECOMMUNICATION INFORMATION:

TELEPHONE: (416) 595-1155

TELEFAX: (416) 595-1163

INFORMATION FOR SEQ ID NO: 14:

SEQUENCE CHARACTERISTICS:

LENGTH: 185 amino acids

TYPE: amino acid

STRANDEDNESS: single

TOPOLOGY: linear

US-08-278-091-14

## Alignment Scores:

Pred. No.:	9.02	Length:	185
Score:	8.00	Matches:	8
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	1.68*	Indels:	0
DB:	1	Gaps:	0

US-10-020-441-1 (1-1477) x US-08-278-091-14 (1-185)

QY 1325 CCAGTGACAGTGTGACCTCTC 1348  
|||||  
Db 138 ProGlyAspSerGlyGlyProLeu 145

## RESULT 4

US-08-483-859-14

; Sequence 14, Application US/08483859  
; Patent No. 5656436

GENERAL INFORMATION:

APPLICANT: LOOSMORE, Sheena M.

APPLICANT: YANG, Yan-Ping

APPLICANT: CHONG, Pele

APPLICANT: COMEN, Raymond P.

APPLICANT: KLEIN, Michel H.

TITLE OF INVENTION: Analog of Haemophilus Hinf7 Protein with

NUMBER OF SEQUENCES: 23

REDUCED PROTEASE ACTIVITY

CORRESPONDENCE ADDRESS:

ADDRESSEE: Sim &amp; McBurney

STREET: Suite 701, 330 University Avenue

CITY: Toronto

STATE: Ontario

COUNTRY: Canada

ZIP: M5G 1R7

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/483,859

FILING DATE: 07-JUN-1995

CLASSIFICATION: 435

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/296,149

FILING DATE: 26-AUG-1994

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/278,091

FILING DATE: 21-JUL-1994

ATTORNEY/AGENT INFORMATION:

NAME: Stewart, Michael I.

REGISTRATION NUMBER: 24,973

REFERENCE/DOCKET NUMBER: 1038-495 MIS:v9

TELECOMMUNICATION INFORMATION:

TELEPHONE: (416) 595-1155

TELEFAX: (416) 595-1163

INFORMATION FOR SEQ ID NO: 14:

SEQUENCE CHARACTERISTICS:

LENGTH: 185 amino acids

TYPE: amino acid

STRANDEDNESS: single

TOPOLOGY: linear

US-08-483-859-14

## Alignment Scores:

Pred. No.:	9.02	Length:	185
Score:	8.00	Matches:	8
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	1.68*	Indels:	0
DB:	1	Gaps:	0

US-10-020-441-1 (1-1477) x US-08-483-859-14 (1-185)

QY 1325 CCAGTGACAGTGTGACCTCTC 1348

Db 138 ProGlyAspSerGlyGlyProLeu 145

## RESULT 5

US-08-472-173-14

; Sequence 14, Application US/08472173  
; Patent No. 5655353

GENERAL INFORMATION:

APPLICANT: LOOSMORE, Sheena M  
APPLICANT: YANG, Yan-Ping  
APPLICANT: CHONG, Pele  
APPLICANT: OOMEN, Raymond P.  
APPLICANT: KLEIN, Michel H.  
TITLE OF INVENTION: Analog of Haemophilus Hln47 Protein with  
REDUCED PROTEASE ACTIVITY  
NUMBER OF SEQUENCES: 23  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Sim & McBurney  
STREET: Suite 701, 330 University Avenue  
CITY: Toronto  
STATE: Ontario  
COUNTRY: Canada  
ZIP: M5G 1R7  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/472,173  
FILING DATE: 07-JUN-1995  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/296,149  
FILING DATE: 26-AUG-1994  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/278,091  
FILING DATE: 21-JUL-1994  
ATTORNEY/AGENT INFORMATION:  
NAME: Stewart, Michael I.  
REGISTRATION NUMBER: 24,973  
REFERENCE/DOCKET NUMBER: 1038-493 MIS:v9  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (416) 595-1155  
TELEFAX: (416) 595-1163  
INFORMATION FOR SEQ ID NO: 14:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 185 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
US-08-472-173-14  
Alignment Scores:  
Pred. No.: 9.02 Length: 185  
Score: 8.00 Matches: 8  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 1.68% Indels: 0  
DB: 1 Gaps: 0  
US-10-020-441-1 (1-1477) x US-08-472-173-14 (1-185)  
OY 1325 CCAGTGACAGTGTGGACCTCTC 1348  
|||||  
Db 138 ProGlyAsperGlyGlyProLeu 145  
RESULT 6  
US-08-487-167-14  
Sequence 14, Application US/08487167  
Patent No. 5869302  
GENERAL INFORMATION:  
APPLICANT: LOOSMORE, Sheena M.  
APPLICANT: YANG, Yan-Ping  
APPLICANT: CHONG, Pele  
APPLICANT: OOMEN, Raymond P.  
APPLICANT: KLEIN, Michel H.  
TITLE OF INVENTION: Analog of Haemophilus Hln47 Protein with  
REDUCED PROTEASE ACTIVITY  
NUMBER OF SEQUENCES: 23  
CORRESPONDENCE ADDRESS:

ADDRESSEE: Sim & McBurney  
STREET: Suite 701, 330 University Avenue  
CITY: Toronto  
STATE: Ontario  
COUNTRY: Canada  
ZIP: M5G 1R7  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/487,167  
FILING DATE: 07-JUN-1995  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/296,149  
FILING DATE: 26-AUG-1994  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/278,091  
FILING DATE: 21-JUL-1994  
ATTORNEY/AGENT INFORMATION:  
NAME: Stewart, Michael I.  
REGISTRATION NUMBER: 24,973  
REFERENCE/DOCKET NUMBER: 1038-508 MIS:v9  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (416) 595-1155  
TELEFAX: (416) 595-1163  
INFORMATION FOR SEQ ID NO: 14:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 185 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
US-08-487-167-14  
Alignment Scores:  
Pred. No.: 9.02 Length: 185  
Score: 8.00 Matches: 8  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 1.68% Indels: 0  
DB: 2 Gaps: 0  
US-10-020-441-1 (1-1477) x US-08-487-167-14 (1-185)  
OY 1325 CCAGTGACAGTGTGGACCTCTC 1348  
|||||  
Db 138 ProGlyAsperGlyGlyProLeu 145  
RESULT 7  
US-08-482-816-14  
Sequence 14, Application US/08482816  
Patent No. 5935573  
GENERAL INFORMATION:  
APPLICANT: LOOSMORE, Sheena M  
APPLICANT: YANG, Yan-Ping  
APPLICANT: CHONG, Pele  
APPLICANT: OOMEN, Raymond P.  
APPLICANT: KLEIN, Michel H.  
TITLE OF INVENTION: Analog of Haemophilus Hln47 Protein with  
REDUCED PROTEASE ACTIVITY  
NUMBER OF SEQUENCES: 23  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Sim & McBurney  
STREET: Suite 701, 330 University Avenue  
CITY: Toronto  
STATE: Ontario  
COUNTRY: Canada  
ZIP: M5G 1R7  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/482,816  
FILING DATE: 07-JUN-1995  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/296,149  
FILING DATE: 26-AUG-1994  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/278,091  
FILING DATE: 21-JUL-1994  
ATTORNEY/AGENT INFORMATION:  
NAME: Stewart, Michael I  
REGISTRATION NUMBER: 24,973  
REFERENCE/DOCKET NUMBER: 1038-494 MIS.vg  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (416) 595-1155  
TELEFAX: (416) 595-1163  
INFORMATION FOR SEQ ID NO: 14:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 185 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
US-08-482-816-14

Alignment Scores:  
Pred. No.: 9.02 Length: 185  
Score: 8.00 Matches: 8  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 1.68% Indels: 0  
DB: 2 Gaps: 0

US-10-020-441-1 (1-1477) x US-08-482-816-14 (1-185)  
QY 1325 CCAGGTGACAGTGTGACCTCTC 1348  
|||||  
DB 138 ProGlyAspserGlyGlyProLeu 145

RESULT 8  
US-08-296-149-14  
Sequence 14, Application US/08296149  
Patent No. 5939297  
GENERAL INFORMATION:  
APPLICANT: LOOSMORE, Sheena M  
APPLICANT: YANG, Yan-Ping  
APPLICANT: CHONG, Pele  
APPLICANT: OOMEN, Raymond P.  
APPLICANT: KLEIN, Michael H.  
TITLE OF INVENTION: Analog of Haemophilus Hinf7 Protein with  
TITLE OF INVENTION: Reduced Protease Activity  
NUMBER OF SEQUENCES: 23  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Slim & McBurney  
STREET: Suite 701, 330 University Avenue  
CITY: Toronto  
STATE: Ontario  
COUNTRY: Canada  
ZIP: MSG 1R7  
COMPUTER READABLE FORM:  
MEDIUM TYPE: floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/296,149  
FILING DATE:  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (416) 595-1155  
TELEFAX: (416) 595-1163  
INFORMATION FOR SEQ ID NO: 14:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 185 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
US-08-296-149-14

TELECOMMUNICATION INFORMATION:  
TELEPHONE: (416) 595-1155  
TELEFAX: (416) 595-1163  
INFORMATION FOR SEQ ID NO: 14:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 185 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
US-08-296-149-14

Alignment Scores:  
Pred. No.: 9.02 Length: 185  
Score: 8.00 Matches: 8  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 1.68% Indels: 0  
DB: 2 Gaps: 0

US-10-020-441-1 (1-1477) x US-08-296-149-14 (1-185)  
QY 1325 CCAGGTGACAGTGTGACCTCTC 1348  
|||||  
DB 138 ProGlyAspserGlyGlyProLeu 145

RESULT 9  
US-08-801-499-14  
Sequence 14, Application US/08801499  
Patent No. 5962430  
GENERAL INFORMATION:  
APPLICANT: LOOSMORE, Sheena M  
APPLICANT: YANG, Yan-Ping  
APPLICANT: CHONG, Pele  
APPLICANT: OOMEN, Raymond P.  
APPLICANT: KLEIN, Michael H.  
TITLE OF INVENTION: Analog of Haemophilus Hinf7 Protein with Reduced Protease A  
TITLE OF INVENTION: Reduced Protease Activity  
NUMBER OF SEQUENCES: 23  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Slim & McBurney  
STREET: 6th Floor, 330 University Avenue  
CITY: Toronto  
STATE: Ontario  
COUNTRY: Canada  
ZIP: MSG 1R7  
COMPUTER READABLE FORM:  
MEDIUM TYPE: floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/801,499  
FILING DATE:  
CLASSIFICATION: 514  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/482,816  
FILING DATE: 07-JUN-1995  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/296,149  
FILING DATE: 26-AUG-1994  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/278,091  
FILING DATE: 21-JUL-1994  
ATTORNEY/AGENT INFORMATION:  
NAME: Stewart, Michael I  
REGISTRATION NUMBER: 24,973  
REFERENCE/DOCKET NUMBER: 1038-671 MIS.jb  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (416) 595-1155  
TELEFAX: (416) 595-1163  
INFORMATION FOR SEQ ID NO: 14:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 185 amino acids  
TYPE: amino acid

```
; STRANDEDNESS: single
; TOPOLOGY: linear
US-08-801-499-14

Alignment Scores:
Pred. No.: 9.02 Length: 185
Score: 8.00 Matches: 8
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 1.68% Indels: 0
DB: 2 Gaps: 0

US-10-020-441-1 (1-1477) x US-08-801-499-14 (1-185)
QY 1325 CCAGGTGACAGTGTGACCTCTC 1348
Db 138 ProGlyaspSerGlyGlyProLeu 145

RESULT 10
US-08-615-271-14
; Sequence 14, Application US/08615271
; Patent No. 5981503
; GENERAL INFORMATION:
; APPLICANT: LOOSMORE, Sheena M.
; APPLICANT: YANG, Yan-Ping
; APPLICANT: CHONG, Pele
; APPLICANT: OOMEN, Raymond P.
; APPLICANT: KLEIN, Michel H.
; TITLE OF INVENTION: ANALOG OF HAEMOPHILUS HIN47 WITH REDUCED
; NUMBER OF SEQUENCES: 23
; CURRENT APPLICATION DATA:
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sim & McBurney
; STREET: 6th Floor, 330 University Avenue
; CITY: Toronto
; STATE: Ontario
; COUNTRY: Canada
; ZIP: M5G 1R7
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/615,271
; FILING DATE: 20-JUN-1996
; CLASSIFICATION: 514
; ATTORNEY/AGENT INFORMATION:
; NAME: Stewart, Michael I.
; REGISTRATION NUMBER: 24,973
; REFERENCE/DOCKET INFORMATION:
; TELECOMMUNICATION INFORMATION: 1038-580
; TELEPHONE: (416) 595-1155
; TELEFAX: (416) 595-1163
; INFORMATION FOR SEQ ID NO: 14:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 185 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
US-08-615-271-14

Alignment Scores:
Pred. No.: 9.02 Length: 185
Score: 8.00 Matches: 8
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 1.68% Indels: 0
DB: 2 Gaps: 0

US-10-020-441-1 (1-1477) x US-08-615-271-14 (1-185)
QY 1325 CCAGGTGACAGTGTGACCTCTC 1348
```

```
Db 138 ProGlyaspSerGlyGlyProLeu 145

RESULT 11
US-09-074-660-14
; Sequence 14, Application US/09074660
; Patent No. 6020183
; GENERAL INFORMATION:
; APPLICANT: LOOSMORE, Sheena M.
; APPLICANT: YANG, Yan-Ping
; APPLICANT: CHONG, Pele
; APPLICANT: OOMEN, Raymond P.
; APPLICANT: KLEIN, Michel H.
; TITLE OF INVENTION: Analog of Haemophilus Hin47 Protein with
; REDUCED PROTEASE ACTIVITY
; NUMBER OF SEQUENCES: 23
; CURRENT APPLICATION DATA:
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sim & McBurney
; STREET: 6th Floor, 330 University Avenue
; CITY: Toronto
; STATE: Ontario
; COUNTRY: Canada
; ZIP: M5G 1R7
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/074,660
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/487,167
; FILING DATE: 07-JUN-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/296,149
; FILING DATE: 26-AUG-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/278,091
; FILING DATE: 21-JUL-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Stewart, Michael I.
; REGISTRATION NUMBER: 24,973
; REFERENCE/DOCKET INFORMATION:
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (416) 595-1155
; TELEFAX: (416) 595-1163
; INFORMATION FOR SEQ ID NO: 14:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 185 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
US-09-074-660-14

Alignment Scores:
Pred. No.: 9.02 Length: 185
Score: 8.00 Matches: 8
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 1.68% Indels: 0
DB: 3 Gaps: 0

US-10-020-441-1 (1-1477) x US-09-074-660-14 (1-185)
QY 1325 CCAGGTGACAGTGTGACCTCTC 1348
Db 138 ProGlyaspSerGlyGlyProLeu 145

RESULT 12
US-09-074-659-14
; Sequence 14, Application US/09074659
```

Patent No. 6025342  
: GENERAL INFORMATION:  
: APPLICANT: LOOSMORE, Sheena M.  
: APPLICANT: YANG, Yan-Ping  
: APPLICANT: CHONG, Pele  
: APPLICANT: OOMEN, Raymond P.  
: APPLICANT: KLEIN, Michel H.  
: TITLE OF INVENTION: Analog of Hemophilus Hinf7 protein with  
: reduced protease activity  
: NUMBER OF SEQUENCES: 23  
: CORRESPONDENCE ADDRESS:  
: ADDRESSEE: Sim & McBurney  
: STREET: 6th floor, 330 University Avenue  
: CITY: Toronto  
: STATE: Ontario  
: COUNTRY: Canada  
: ZIP: M5G 1R7  
: COMPUTER READABLE FORM:  
: MEDIUM TYPE: floppy disk  
: COMPUTER: IBM PC compatible  
: OPERATING SYSTEM: PC-DOS/MS-DOS  
: SOFTWARE: Patentin Release #1.0, Version #1.25  
: CURRENT APPLICATION DATA:  
: APPLICATION NUMBER: US/09/074,659  
: FILING DATE:  
: CLASSIFICATION:  
: PRIOR APPLICATION DATA:  
: APPLICATION NUMBER: US 08/487,167  
: FILING DATE: 07-JUN-1995  
: PRIOR APPLICATION DATA:  
: APPLICATION NUMBER: US 08/296,149  
: FILING DATE: 26-AUG-1994  
: PRIOR APPLICATION DATA:  
: APPLICATION NUMBER: US 08/278,091  
: FILING DATE: 21-JUL-1994  
: ATTORNEY/AGENT INFORMATION:  
: NAME: Stewart, Michael I.  
: REGISTRATION NUMBER: 24,973  
: REFERENCE/DOCKET NUMBER: 1038-730 MIS:jb  
: TELECOMMUNICATION INFORMATION:  
: TELEPHONE: (416) 595-1155  
: TELEFAX: (416) 595-1163  
: INFORMATION FOR SEQ ID NO: 14:  
: SEQUENCE CHARACTERISTICS:  
: LENGTH: 185 amino acids  
: TYPE: amino acid  
: STRANDEDNESS: single  
: TOPOLOGY: linear  
: US-09-074-659-14

Alignment Scores:  
Pred. No.: 9.02 Length: 185  
Score: 8.00 Matches: 8  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 1.68% Indels: 0  
DB: 3 Gaps: 0

US-10-020-441-1 (1-1477) x US-09-074-659-14 (1-185)  
OY 1325 CCAGGTGACAGTGTGACCTCTC 1348  
DB 138 ProGlyAspserGlyGlyProLeu 145

RESULT 13  
US-09-106-468-14  
: Sequence 14, Application US/09106468  
: Patent No. 6114125  
: GENERAL INFORMATION:  
: APPLICANT: LOOSMORE, Sheena M.  
: APPLICANT: YANG, Yan-Ping  
: APPLICANT: CHONG, Pele  
: APPLICANT: OOMEN, Raymond P.

APPLICANT: KLEIN, Michel H.  
: TITLE OF INVENTION: ANALOG OF HAEMOPHILUS HINF7 WITH REDUCED  
: PROTEASE ACTIVITY  
: NUMBER OF SEQUENCES: 23  
: CORRESPONDENCE ADDRESS:  
: ADDRESSEE: Sim & McBurney  
: STREET: 6th floor, 330 University Avenue  
: CITY: Toronto  
: STATE: Ontario  
: COUNTRY: Canada  
: ZIP: M5G 1R7  
: COMPUTER READABLE FORM:  
: MEDIUM TYPE: floppy disk  
: COMPUTER: IBM PC compatible  
: OPERATING SYSTEM: PC-DOS/MS-DOS  
: SOFTWARE: Patentin Release #1.0, Version #1.30  
: CURRENT APPLICATION DATA:  
: APPLICATION NUMBER: US/09/106,468  
: FILING DATE:  
: CLASSIFICATION:  
: PRIOR APPLICATION DATA:  
: APPLICATION NUMBER: US 08/615,271  
: FILING DATE: 20-JUN-1996  
: CLASSIFICATION:  
: ATTORNEY/AGENT INFORMATION:  
: NAME: Stewart, Michael I.  
: REGISTRATION NUMBER: 24,973  
: REFERENCE/DOCKET NUMBER: 1038-825  
: TELECOMMUNICATION INFORMATION:  
: TELEPHONE: (416) 595-1155  
: TELEFAX: (416) 595-1163  
: INFORMATION FOR SEQ ID NO: 14:  
: SEQUENCE CHARACTERISTICS:  
: LENGTH: 185 amino acids  
: TYPE: amino acid  
: STRANDEDNESS: single  
: TOPOLOGY: linear  
: US-09-106-468-14

Alignment Scores:  
Pred. No.: 9.02 Length: 185  
Score: 8.00 Matches: 8  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 1.68% Indels: 0  
DB: 3 Gaps: 0

US-10-020-441-1 (1-1477) x US-09-106-468-14 (1-185)  
OY 1325 CCAGGTGACAGTGTGACCTCTC 1348  
DB 138 ProGlyAspserGlyGlyProLeu 145

RESULT 14  
US-09-106-466A-14  
: Sequence 14, Application US/09106466A  
: Patent No. 6147057  
: GENERAL INFORMATION:  
: APPLICANT: LOOSMORE, Sheena M.  
: APPLICANT: YANG, Yan-Ping  
: APPLICANT: CHONG, Pele  
: APPLICANT: OOMEN, Raymond P.  
: APPLICANT: KLEIN, Michel H.  
: TITLE OF INVENTION: ANALOG OF HAEMOPHILUS HINF7 WITH REDUCED  
: PROTEASE ACTIVITY  
: NUMBER OF SEQUENCES: 23  
: CORRESPONDENCE ADDRESS:  
: ADDRESSEE: Sim & McBurney  
: STREET: 6th floor, 330 University Avenue  
: CITY: Toronto  
: STATE: Ontario  
: COUNTRY: Canada  
: ZIP: M5G 1R7

```
;
; COMPUTER READABLE FORM:
;
; MEDIUM TYPE: Floppy disk
;
; COMPUTER: IBM PC compatible
;
; OPERATING SYSTEM: PC-DOS/MS-DOS
;
; SOFTWARE: Patentin Release #1.0, Version #1.30
;
; CURRENT APPLICATION DATA:
;
; APPLICATION NUMBER: US/09/106,466A
;
; FILING DATE:
;
; CLASSIFICATION: 514514
;
; PRIOR APPLICATION DATA:
;
; APPLICATION NUMBER: US 08/615,271
;
; FILING DATE: 20-JUN-1996
;
; CLASSIFICATION: 514514
;
; ATTORNEY/AGENT INFORMATION:
;
; NAME: Stewart, Michael I
;
; REGISTRATION NUMBER: 24,973
;
; REFERENCE/DOCKET NUMBER: 1038-826
;
; TELECOMMUNICATION INFORMATION:
;
; TELEPHONE: (416) 595-1155
;
; TELEFAX: (416) 595-1163
;
; INFORMATION FOR SEQ ID NO: 14:
;
; SEQUENCE CHARACTERISTICS:
;
; LENGTH: 185 amino acids
;
; TYPE: amino acid
;
; STRANDEDNESS: single
;
; TOPOLOGY: linear
;
; US-09-106-466A-14
;
;
; Alignment Scores:
;
; Pred. No.: 9.02 Length: 185
;
; Score: 8.00 Matches: 8
;
; Percent Similarity: 100.00% Conservative: 0
;
; Best Local Similarity: 100.00% Mismatches: 0
;
; Query Match: 1.68% Indels: 0
;
; Gaps: 0
;
; DB: 4
;
; US-10-020-441-1 (1-1477) x US-09-106-466A-14 (1-185)
;
;
; QY 1325 CCAGGTGACAGTGGTGACCTCTC 1348
;
; Db 138 ProGlyAspserGlyGlyProLeu 145
;
;
; RESULT 15
;
; US-09-106-467-14
;
; Sequence 14, Application US/09106467
;
; Patent No. 6135380
;
; GENERAL INFORMATION:
;
; APPLICANT: LOOSMORE, Sheena M.
;
; APPLICANT: YANG, Yan-Ping
;
; APPLICANT: CHONG, Pele
;
; APPLICANT: COHEN, Raymond P.
;
; APPLICANT: KLEIN, Michel H.
;
; TITLE OF INVENTION: ANALOG OF HAEMOPHILUS HIN47 WITH REDUCED
;
; NUMBER OF SEQUENCES: 23
;
; CORRESPONDENCE ADDRESS:
;
; ADDRESSEE: Sim & McBurney
;
; STREET: 6th Floor, 330 University Avenue
;
; CITY: Toronto
;
; STATE: Ontario
;
; COUNTRY: Canada
;
; ZIP: M5G 1R7
;
; COMPUTER READABLE FORM:
;
; MEDIUM TYPE: Floppy disk
;
; COMPUTER: IBM PC compatible
;
; OPERATING SYSTEM: PC-DOS/MS-DOS
;
; SOFTWARE: Patentin Release #1.0, Version #1.30
;
; CURRENT APPLICATION DATA:
;
; APPLICATION NUMBER: US/09/106,467
;
; FILING DATE:
;
; CLASSIFICATION:
;
; PRIOR APPLICATION DATA:
;
; APPLICATION NUMBER: US 08/615,271
```

```
;
; FILING DATE: 20-JUN-1996
;
; CLASSIFICATION:
;
; ATTORNEY/AGENT INFORMATION:
;
; NAME: Stewart, Michael I
;
; REGISTRATION NUMBER: 24,973
;
; REFERENCE/DOCKET NUMBER: 1038-824
;
; TELECOMMUNICATION INFORMATION:
;
; TELEPHONE: (416) 595-1155
;
; TELEFAX: (416) 595-1163
;
; INFORMATION FOR SEQ ID NO: 14:
;
; SEQUENCE CHARACTERISTICS:
;
; LENGTH: 185 amino acids
;
; TYPE: amino acid
;
; STRANDEDNESS: single
;
; TOPOLOGY: linear
;
; US-09-106-467-14
;
;
; Alignment Scores:
;
; Pred. No.: 9.02 Length: 185
;
; Score: 8.00 Matches: 8
;
; Percent Similarity: 100.00% Conservative: 0
;
; Best Local Similarity: 100.00% Mismatches: 0
;
; Query Match: 1.68% Indels: 0
;
; Gaps: 0
;
; DB: 4
;
; US-10-020-441-1 (1-1477) x US-09-106-467-14 (1-185)
;
;
; QY 1325 CCAGGTGACAGTGGTGACCTCTC 1348
;
; Db 138 ProGlyAspserGlyGlyProLeu 145
;
;
; RESULT 16
;
; 5514590-10
;
; Patent No. 5514590
;
; APPLICANT: GARVIN, ROBERT T.; MALEK, LAWRENCE T.; JAMES, ERIC
;
; TITLE OF INVENTION: EXPRESSION SYSTEM COMPRISING DNA
;
; ENCODING THE SIGNAL PEPTIDE OF PROTEASE B FROM STREPTOMYCES
;
; GRISEUS
;
; NUMBER OF SEQUENCES: 12
;
; CURRENT APPLICATION DATA:
;
; APPLICATION NUMBER: US/08/203,644
;
; FILING DATE: 01-MAR-1994
;
; PRIOR APPLICATION DATA:
;
; APPLICATION NUMBER: 66,938
;
; FILING DATE: 25-MAY-1993
;
; APPLICATION NUMBER: 844,937
;
; FILING DATE: 04-MAR-1992
;
; APPLICATION NUMBER: 221,346
;
; FILING DATE: 18-JUL-1988
;
; APPLICATION NUMBER: 795,331
;
; FILING DATE: 06-NOV-1985
;
; SEQ ID NO:10:
;
; LENGTH: 185
;
; 5514590-10
;
;
; Alignment Scores:
;
; Pred. No.: 9.02 Length: 185
;
; Score: 8.00 Matches: 8
;
; Percent Similarity: 100.00% Conservative: 0
;
; Best Local Similarity: 100.00% Mismatches: 0
;
; Query Match: 1.68% Indels: 0
;
; Gaps: 0
;
; DB: 6
;
; US-10-020-441-1 (1-1477) x 5514590-10 (1-185)
;
;
; QY 1325 CCAGGTGACAGTGGTGACCTCTC 1348
;
; Db 138 ProGlyAspserGlyGlyProLeu 145
;
;
; RESULT 17
;
; US-07-956-848A-41
;
; Sequence 41, Application US/07956848A
;
; Patent No. 5447914
```

GENERAL INFORMATION:  
APPLICANT: Travis, James  
APPLICANT: Shafer, William M.  
APPLICANT: Bangalore, Neelash  
APPLICANT: Pohl, Jan  
TITLE OF INVENTION: Antimicrobial Peptides  
NUMBER OF SEQUENCES: 59  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Greenlee and Winner, P.C.  
STREET: 5370 Manhattan Circle, Suite 201  
CITY: Boulder  
STATE: Colorado  
COUNTRY: USA  
ZIP: 80303  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/07/956,848A  
FILING DATE: 02-OCT-1992  
CLASSIFICATION: 514  
ATTORNEY/AGENT INFORMATION:  
NAME: Feider, Donna M.  
REGISTRATION NUMBER: 33,878  
REFERENCE/DOCKET NUMBER: 3-90B  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (303) 499-8080  
TELEFAX: (303) 499-8089  
TELEX: 49617824  
INFORMATION FOR SEQ ID NO: 41:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 223 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
US-07-956-848A-41

Alignment Scores:  
Pred. No.: 8.9 Length: 223  
Score: 8.00 Matches: 8  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 1.68% Indels: 0  
DB: 1 Gaps: 0

US-10-020-441-1 (1-1477) x US-07-956-848A-41 (1-223)  
Qy 1328 GGTGACAGTGTGACCTCTCCTC 1351  
|||||  
Db 179 GLYASPSERGLYGLYPROLEUEN 186

RESULT 18  
US-08-485-455D-17  
Sequence 17, Application US/08485455D  
Patent No. 5712143  
GENERAL INFORMATION:  
APPLICANT: Grileve, Robert B.  
APPLICANT: Rushlow, Keith E.  
APPLICANT: Wu Hunter, Shirley  
APPLICANT: Frank, Glenn R.  
APPLICANT: Stiegler, Gary L.  
TITLE OF INVENTION: FLEA PROTEASE PROTEINS, NUCLEIC ACID  
NUMBER OF SEQUENCES: 79  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Sheridan Ross P.C.  
STREET: 1700 Lincoln Street, Suite 3500  
CITY: Denver  
STATE: Colorado  
COUNTRY: USA

ZIP: 80203  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/485,455D  
FILING DATE: 07-JUN-1995  
CLASSIFICATION: 424  
ATTORNEY/AGENT INFORMATION:  
NAME: Connell, Gary J.  
REGISTRATION NUMBER: 32,020  
REFERENCE/DOCKET NUMBER: 2618-25-C1-3  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (303) 863-9700  
TELEFAX: (303) 863-0223  
INFORMATION FOR SEQ ID NO: 17:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 223 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-485-455D-17

Alignment Scores:  
Pred. No.: 8.9 Length: 223  
Score: 8.00 Matches: 8  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 1.68% Indels: 0  
DB: 1 Gaps: 0

US-10-020-441-1 (1-1477) x US-08-485-455D-17 (1-223)  
Qy 1328 GGTGACAGTGTGACCTCTCCTC 1351  
|||||  
Db 177 GLYASPSERGLYGLYPROLEUEN 184

RESULT 19  
US-08-471-956-41  
Sequence 41, Application US/08471956  
Patent No. 5798336  
GENERAL INFORMATION:  
APPLICANT: Travis, James  
APPLICANT: Shafer, William M.  
APPLICANT: Bangalore, Neelash  
APPLICANT: Pohl, Jan  
TITLE OF INVENTION: Antimicrobial Peptides  
NUMBER OF SEQUENCES: 59  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Greenlee and Winner, P.C.  
STREET: 5370 Manhattan Circle, Suite 201  
CITY: Boulder  
STATE: Colorado  
COUNTRY: USA  
ZIP: 80303  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/471,956  
FILING DATE: 06-JUN-1995  
CLASSIFICATION: 514  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/956,848  
FILING DATE: 02-OCT-1992  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/541,635  
FILING DATE: 21-JUN-1990  
ATTORNEY/AGENT INFORMATION:

NAME: Ferber, Donna M.  
REGISTRATION NUMBER: 33,878  
REFERENCE/DOCKET NUMBER: 3-90D  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (303) 499-8080  
TELEFAX: (303) 499-8089  
TELEX: 49617824  
INFORMATION FOR SEQ ID NO: 41:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 223 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
US-08-471-956-41

Alignment Scores:  
Pred. No.: 8.9 Length: 223  
Score: 8.00 Matches: 8  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 1.68% Indels: 0  
DB: 1 Gaps: 0

US-10-020-441-1 (1-1477) x US-08-471-956-41 (1-223)  
QY 1328 GGTGACAGTGTGGACCTCTCCTC 1351  
|||||  
Db 179 GlyaspserglyglyProleuLeu 186

RESULT 20  
US-08-482-130C-17  
Sequence 17, Application US/08482130C  
Patent No. 5962257  
GENERAL INFORMATION:  
APPLICANT: Grieve, Robert B.  
APPLICANT: Rushlow, Keith E.  
APPLICANT: Wu Hunter, Shirley  
APPLICANT: Frank, Glenn R.  
APPLICANT: Stiegler, Gary L.  
TITLE OF INVENTION: FLEA PROTEASE PROTEINS, NUCLEIC ACID  
TITLE OF INVENTION: MOLECULES, AND USES THEREOF  
NUMBER OF SEQUENCES: 79  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Sheridan Ross P.C.  
STREET: 1700 Lincoln Street, Suite 3500  
CITY: Denver  
STATE: Colorado  
COUNTRY: USA  
ZIP: 80203  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/482,130C  
FILING DATE: 07-JUN-1995  
CLASSIFICATION: 424  
ATTORNEY/AGENT INFORMATION:  
NAME: Connell, Gary J.  
REGISTRATION NUMBER: 32,020  
REFERENCE/DOCKET NUMBER: 2618-25-C1-1  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (303) 863-9700  
TELEFAX: (303) 863-0223  
INFORMATION FOR SEQ ID NO: 17:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 223 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-482-130C-17

Alignment Scores:  
Pred. No.: 8.9 Length: 223  
Score: 8.00 Matches: 8  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 1.68% Indels: 0  
DB: 2 Gaps: 0

US-10-020-441-1 (1-1477) x US-08-482-130C-17 (1-223)  
QY 1328 GGTGACAGTGTGGACCTCTCCTC 1351  
|||||  
Db 177 GlyaspserglyglyProleuLeu 184

RESULT 21  
US-08-484-211C-17  
Sequence 17, Application US/08484211C  
Patent No. 5972645  
GENERAL INFORMATION:  
APPLICANT: Grieve, Robert B.  
APPLICANT: Rushlow, Keith E.  
APPLICANT: Wu Hunter, Shirley  
APPLICANT: Frank, Glenn R.  
APPLICANT: Stiegler, Gary L.  
TITLE OF INVENTION: FLEA PROTEASE PROTEINS, NUCLEIC ACID  
TITLE OF INVENTION: MOLECULES, AND USES THEREOF  
NUMBER OF SEQUENCES: 79  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Sheridan Ross P.C.  
STREET: 1700 Lincoln Street, Suite 3500  
CITY: Denver  
STATE: Colorado  
COUNTRY: USA  
ZIP: 80203  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/484,211C  
FILING DATE: 07-JUN-1995  
CLASSIFICATION: 424  
ATTORNEY/AGENT INFORMATION:  
NAME: Connell, Gary J.  
REGISTRATION NUMBER: 32,020  
REFERENCE/DOCKET NUMBER: 2618-25-C1  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (303) 863-9700  
TELEFAX: (303) 863-0223  
INFORMATION FOR SEQ ID NO: 17:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 223 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-484-211C-17

Alignment Scores:  
Pred. No.: 8.9 Length: 223  
Score: 8.00 Matches: 8  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 1.68% Indels: 0  
DB: 2 Gaps: 0

US-10-020-441-1 (1-1477) x US-08-484-211C-17 (1-223)  
QY 1328 GGTGACAGTGTGGACCTCTCCTC 1351  
|||||  
Db 177 GlyaspserglyglyProleuLeu 184



RESULT 22  
US-08-817-795-17  
; Sequence 17, Application US/08817795  
; Patent No. 6139840  
; GENERAL INFORMATION:  
; APPLICANT: Grievé, Robert B.  
; APPLICANT: Rushlow, Keith E.  
; APPLICANT: Hunter, Shirley Wu  
; APPLICANT: Frank, Glenn R.  
; APPLICANT: Heath, Andrew W.  
; APPLICANT: Yamaka, Miles Yamanaka  
; APPLICANT: Arfsten, Ann  
; APPLICANT: Dale, Beverly  
; APPLICANT: Stiegler, Gary  
; TITLE OF INVENTION: USE OF PROTEASE INHIBITORS AND  
; TITLE OF INVENTION: PROTEASE VACCINES TO PROTECT ANIMALS FROM FLEA  
; TITLE OF INVENTION: INFESTATION, AND FLEA PROTEASE PROTEINS, NUCLEIC ACID  
; TITLE OF INVENTION: MOLECULES, AND USES THEREOF  
; NUMBER OF SEQUENCES: 119  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Sheridan Ross & McIntosh  
; STREET: 1700 Lincoln Street, Suite 3500  
; CITY: Denver  
; STATE: Colorado  
; COUNTRY: USA  
; ZIP: 80203  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent In Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/817,795  
; FILING DATE:  
; CLASSIFICATION:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: PCT/US95/14442  
; FILING DATE:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Gary J. Connell  
; REGISTRATION NUMBER: 32,020  
; REFERENCE/DOCKET NUMBER:  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (303) 863-9700  
; TELEFAX: (303) 863-0223  
; INFORMATION FOR SEQ ID NO: 17:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 223 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-08-817-795-17

Alignment Scores:  
Pred. No.: 8.9 Length: 223  
Score: 8.00 Matches: 8  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 1.68% Indels: 0  
DB: 4 Gaps: 0

US-10-020-441-1 (1-1477) x US-08-817-795-17 (1-223)

QY 1328 GGTGACAGTGTGACCTCTCCTC 1351  
Db 177 GlyAspSerGlyGlyProLeuLeu 184

RESULT 23  
US-08-485-443B-17  
; Sequence 17, Application US/08485443B  
; Patent No. 6148870  
; GENERAL INFORMATION:  
; APPLICANT: Grievé, Robert B.

APPLICANT: Rushlow, Keith E.  
APPLICANT: Hunter, Shirley Wu  
APPLICANT: Frank, Glenn R.  
APPLICANT: Stiegler, Gary L.  
TITLE OF INVENTION: FLEA PROTEASE PROTEINS, NUCLEIC ACID  
TITLE OF INVENTION: MOLECULES AND USES THEREOF  
NUMBER OF SEQUENCES: 79  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: SHERIDAN ROSS  
STREET: 1700 LINCOLN ST., SUITE 3500  
CITY: DENVER  
STATE: CO  
COUNTRY: USA  
ZIP: 80203  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent In Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/485,443B  
FILING DATE: 07-JUN-1995  
CLASSIFICATION: 424  
ATTORNEY/AGENT INFORMATION:  
NAME: Connell, Gary L.  
REGISTRATION NUMBER: 32,020  
REFERENCE/DOCKET NUMBER: 2618-25-C1-2  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 303 863-9700  
TELEFAX: 303 863-0223  
INFORMATION FOR SEQ ID NO: 17:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 223 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-485-443B-17

Alignment Scores:  
Pred. No.: 8.9 Length: 223  
Score: 8.00 Matches: 8  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 1.68% Indels: 0  
DB: 4 Gaps: 0

US-10-020-441-1 (1-1477) x US-08-485-443B-17 (1-223)

QY 1328 GGTGACAGTGTGACCTCTCCTC 1351  
Db 177 GlyAspSerGlyGlyProLeuLeu 184

RESULT 24  
PCT-US95-14442A-17  
; Sequence 17, Application PC/TUS9514442A  
; GENERAL INFORMATION:  
; APPLICANT: Grievé, Robert B.  
; APPLICANT: Rushlow, Keith E.  
; APPLICANT: Hunter, Shirley Wu  
; APPLICANT: Frank, Glenn R.  
; APPLICANT: Heath, Andrew W.  
; APPLICANT: Yamaka, Miles Yamanaka  
; APPLICANT: Arfsten, Ann  
; APPLICANT: Dale, Beverly  
; APPLICANT: Stiegler, Gary  
; TITLE OF INVENTION: USE OF PROTEASE INHIBITORS AND  
; TITLE OF INVENTION: PROTEASE VACCINES TO PROTECT ANIMALS FROM FLEA  
; TITLE OF INVENTION: INFESTATION, AND FLEA PROTEASE PROTEINS, NUCLEIC ACID  
; TITLE OF INVENTION: MOLECULES, AND USES THEREOF  
; NUMBER OF SEQUENCES: 119  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Sheridan Ross & McIntosh  
; STREET: 1700 Lincoln Street, Suite 3500

CITY: Denver  
STATE: Colorado  
COUNTRY: USA  
ZIP: 80203  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: PCT/US95/14442A  
FILING DATE:  
CLASSIFICATION:  
ATTORNEY/AGENT INFORMATION:  
NAME: Gary J. Connell  
REGISTRATION NUMBER: 32,020  
REFERENCE/DOCKET NUMBER:  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (303) 863-9700  
TELEFAX: (303) 863-0223  
INFORMATION FOR SEQ ID NO: 17:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 223 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
PCT-US95-14442A-17

Alignment Scores:  
Pred. No.: 8.9 Length: 223  
Score: 8.00 Matches: 8  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 1.66% Indels: 0  
DB: 5 Gaps: 0

US-10-020-441-1 (1-1477) x PCT-US95-14442A-17 (1-223)  
QY 1328 GGTGACAGTGGTGACCTCTCCTC 1351  
Db 177 G1Yspserg1g1YProLeu1eu 184

RESULT 25  
US-08-278-091-12  
Sequence 12, Application US/08278091  
Patent No. 5506139  
GENERAL INFORMATION:  
APPLICANT: LOOSMORE, Sheena M  
APPLICANT: YANG, Yan-ping  
APPLICANT: CHONG, Pele  
APPLICANT: OOMEN, Raymond P.  
APPLICANT: KLEIN, Michel H.  
TITLE OF INVENTION: Analog of Haemophilus H1n47 protein with  
NUMBER OF INVENTIONS: Reduced Protease Activity  
NUMBER OF SEQUENCES: 23  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Sim & McBurney  
STREET: Suite 701, 330 University Avenue  
CITY: Toronto  
STATE: Ontario  
COUNTRY: Canada  
ZIP: M5G 1R7  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/278,091  
FILING DATE: 21-JUL-1994  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Stewart, Michael I

REGISTRATION NUMBER: 24,973  
REFERENCE/DOCKET NUMBER: 1038-371  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (416) 595-1155  
TELEFAX: (416) 595-1163  
INFORMATION FOR SEQ ID NO: 12:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 224 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
US-08-278-091-12

Alignment Scores:  
Pred. No.: 8.9 Length: 224  
Score: 8.00 Matches: 8  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 1.66% Indels: 0  
DB: 1 Gaps: 0

US-10-020-441-1 (1-1477) x US-08-278-091-12 (1-224)  
QY 1328 GGTGACAGTGGTGACCTCTCCTC 1351  
Db 180 G1Yspserg1g1YProLeu1eu 187

RESULT 26  
US-08-483-859-12  
Sequence 12, Application US/08483859  
Patent No. 5656436  
GENERAL INFORMATION:  
APPLICANT: LOOSMORE, Sheena M.  
APPLICANT: YANG, Yan-ping  
APPLICANT: CHONG, Pele  
APPLICANT: OOMEN, Raymond P.  
APPLICANT: KLEIN, Michel H.  
TITLE OF INVENTION: Analog of Haemophilus H1n47 Protein with  
NUMBER OF INVENTIONS: Reduced Protease Activity  
NUMBER OF SEQUENCES: 23  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Sim & McBurney  
STREET: Suite 701, 330 University Avenue  
CITY: Toronto  
STATE: Ontario  
COUNTRY: Canada  
ZIP: M5G 1R7  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/483,859  
FILING DATE: 07-JUN-1995  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/296,149  
FILING DATE: 26-AUG-1994  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/278,091  
FILING DATE: 21-JUL-1994  
ATTORNEY/AGENT INFORMATION:  
NAME: Stewart, Michael I.  
REGISTRATION NUMBER: 24,973  
REFERENCE/DOCKET NUMBER: 1038-495 MIS:vg  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (416) 595-1155  
TELEFAX: (416) 595-1163  
INFORMATION FOR SEQ ID NO: 12:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 224 amino acids  
TYPE: amino acid

STRANDEDNESS: single  
TOPOLOGY: linear  
US-08-483-859-12

Alignment Scores:

Pred. No.:	8.9	Length:	224
Score:	8.00	Matches:	8
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	1.68%	Indels:	0
DB:	1	Gaps:	0

US-10-020-441-1 (1-1477) x US-08-483-859-12 (1-224)

OY 1328 GGTGACAGTGTGACCTCTCCTC 1351  
DB 180 GIYAspserGlyGlyProleuLeu 187

RESULT 27  
US-08-472-173-12

Sequence 12, Application US/08472173  
Patent No. 5665353

GENERAL INFORMATION:  
APPLICANT: LOOSMORE, Sheena M  
APPLICANT: YANG, Yan-Ping  
APPLICANT: CHONG, Pele  
APPLICANT: COMEN, Raymond P.  
APPLICANT: KLEIN, Michel H.  
TITLE OF INVENTION: Analog of Haemophilus Hinf7 Protein with  
NUMBER OF SEQUENCES: 23  
REDUCED PROTEASE ACTIVITY

CORRESPONDENCE ADDRESS:  
ADDRESSEE: Sim & McBurney  
STREET: Suite 701, 330 University Avenue  
CITY: Toronto  
STATE: Ontario  
COUNTRY: Canada  
ZIP: M5G 1R7

COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25

CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/472.173  
FILING DATE: 07-JUN-1995  
CLASSIFICATION: 435

PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/296,149  
FILING DATE: 26-AUG-1994  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/278,091  
FILING DATE: 21-JUL-1994  
ATTORNEY/AGENT INFORMATION:  
NAME: Stewart, Michael I.  
REGISTRATION NUMBER: 24,973  
REFERENCE/DOCKET NUMBER: 1038-493 MIS:vg  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (416) 595-1155  
TELEFAX: (416) 595-1153  
INFORMATION FOR SEQ ID NO: 12:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 224 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear

US-08-472-173-12

Alignment Scores:

Pred. No.:	8.9	Length:	224
Score:	8.00	Matches:	8
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	1.68%	Indels:	0
DB:	1	Gaps:	0

Query Match: 1.68% Indels: 0  
DB: 1 Gaps: 0

US-10-020-441-1 (1-1477) x US-08-472-173-12 (1-224)

OY 1328 GGTGACAGTGTGACCTCTCCTC 1351  
DB 180 GIYAspserGlyGlyProleuLeu 187

RESULT 28  
US-08-487-167-12

Sequence 12, Application US/08487167  
Patent No. 5869302

GENERAL INFORMATION:  
APPLICANT: LOOSMORE, Sheena M.  
APPLICANT: YANG, Yan-Ping  
APPLICANT: CHONG, Pele  
APPLICANT: COMEN, Raymond P.  
APPLICANT: KLEIN, Michel H.  
TITLE OF INVENTION: Analog of Haemophilus Hinf7 Protein with  
NUMBER OF SEQUENCES: 23  
REDUCED PROTEASE ACTIVITY

CORRESPONDENCE ADDRESS:  
ADDRESSEE: Sim & McBurney  
STREET: Suite 701, 330 University Avenue  
CITY: Toronto  
STATE: Ontario  
COUNTRY: Canada  
ZIP: M5G 1R7

COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25

CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/487,167  
FILING DATE: 07-JUN-1995  
CLASSIFICATION: 435

PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/296,149  
FILING DATE: 26-AUG-1994  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/278,091  
FILING DATE: 21-JUL-1994  
ATTORNEY/AGENT INFORMATION:  
NAME: Stewart, Michael I.  
REGISTRATION NUMBER: 24,973  
REFERENCE/DOCKET NUMBER: 1038-508 MIS:vg  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (416) 595-1155  
TELEFAX: (416) 595-1153  
INFORMATION FOR SEQ ID NO: 12:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 224 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear

US-08-487-167-12

Alignment Scores:

Pred. No.:	8.9	Length:	224
Score:	8.00	Matches:	8
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	1.68%	Indels:	0
DB:	2	Gaps:	0

US-10-020-441-1 (1-1477) x US-08-487-167-12 (1-224)

OY 1328 GGTGACAGTGTGACCTCTCCTC 1351  
DB 180 GIYAspserGlyGlyProleuLeu 187

```
RESULT 29
US-08-482-816-12
; Sequence 12, Application US/08482816
; Patent No. 5935573
; GENERAL INFORMATION:
; APPLICANT: LOOSMORE, Sheena M
; APPLICANT: CHONG, Pele
; APPLICANT: OOMEN, Raymond P.
; APPLICANT: KLEIN, Michel H.
; TITLE OF INVENTION: Analog of Haemophilus Hln47 Protein with Reduced Protease Act
; NUMBER OF SEQUENCES: 23
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sim & McBurney
; STREET: Suite 701, 330 University Avenue
; CITY: Toronto
; STATE: Ontario
; COUNTRY: Canada
; ZIP: M5G 1R7
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/482,816
; FILING DATE: 07-JUN-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/296,149
; FILING DATE: 26-AUG-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/278,091
; FILING DATE: 21-JUL-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Stewart, Michael I
; REGISTRATION NUMBER: 24,973
; REFERENCE/DOCKET NUMBER: 1038-494 MIS:v9
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (416) 595-1155
; TELEFAX: (416) 595-1163
; INFORMATION FOR SEQ ID NO: 12:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 224 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; US-08-482-816-12

Alignment Scores:
Pred. No.: 8.9 Length: 224
Score: 8.00 Matches: 8
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 1.68% Indels: 0
DB: 2 Gaps: 0

US-10-020-441-1 (1-1477) x US-08-482-816-12 (1-224)
QY 1328 GGTGACAGTGGTGACCTCTCCTC 1351
|||||
Db 180 GYAspergilyglyProleuLeu 187

RESULT 30
US-08-296-149-12
; Sequence 12, Application US/08296149
; Patent No. 5939297
; GENERAL INFORMATION:
; APPLICANT: LOOSMORE, Sheena M
; APPLICANT: CHONG, Pele
; APPLICANT: OOMEN, Raymond P.
; APPLICANT: KLEIN, Michel H.
```

```
; TITLE OF INVENTION: Analog of Haemophilus Hln47 Protein with
; TITLE OF INVENTION: Reduced Protease Activity
; NUMBER OF SEQUENCES: 23
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sim & McBurney
; STREET: Suite 701, 330 University Avenue
; CITY: Toronto
; STATE: Ontario
; COUNTRY: Canada
; ZIP: M5G 1R7
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/296,149
; FILING DATE:
; CLASSIFICATION: 514
; ATTORNEY/AGENT INFORMATION:
; NAME: Stewart, Michael I
; REGISTRATION NUMBER: 24,973
; REFERENCE/DOCKET NUMBER: 1038-390
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (416) 595-1155
; TELEFAX: (416) 595-1163
; INFORMATION FOR SEQ ID NO: 12:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 224 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; US-08-296-149-12

Alignment Scores:
Pred. No.: 8.9 Length: 224
Score: 8.00 Matches: 8
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 1.68% Indels: 0
DB: 2 Gaps: 0

US-10-020-441-1 (1-1477) x US-08-296-149-12 (1-224)
QY 1328 GGTGACAGTGGTGACCTCTCCTC 1351
|||||
Db 180 GYAspergilyglyProleuLeu 187

RESULT 31
US-08-801-499-12
; Sequence 12, Application US/08801499
; Patent No. 5962430
; GENERAL INFORMATION:
; APPLICANT: LOOSMORE, Sheena M
; APPLICANT: CHONG, Pele
; APPLICANT: OOMEN, Raymond P.
; APPLICANT: KLEIN, Michel H.
; TITLE OF INVENTION: Analog of Haemophilus Hln47 Protein with Reduced Protease A
; NUMBER OF SEQUENCES: 23
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sim & McBurney
; STREET: 6th Floor, 330 University Avenue
; CITY: Toronto
; STATE: Ontario
; COUNTRY: Canada
; ZIP: M5G 1R7
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
```

```
APPLICATION NUMBER: US/08/801,499
FILING DATE:
CLASSIFICATION: 514
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/482,816
FILING DATE: 07-JUN-1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/296,149
FILING DATE: 26-AUG-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/278,091
FILING DATE: 21-JUL-1994
ATTORNEY/AGENT INFORMATION:
NAME: Stewart, Michael I
REGISTRATION NUMBER: 24,973
REFERENCE/DOCKET NUMBER: 1038-671 MIS:jfb
TELECOMMUNICATION INFORMATION:
TELEPHONE: (416) 595-1155
TELEFAX: (416) 595-1163
INFORMATION FOR SEQ ID NO: 12:
SEQUENCE CHARACTERISTICS:
LENGTH: 224 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-801-499-12

Alignment Scores:
Pred. No.: 8.9 Length: 224
Score: 8.00 Matches: 8
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 1.68% Indels: 0
DB: 2 Gaps: 0

US-10-020-441-1 (1-1477) x US-08-801-499-12 (1-224)
QY 1328 GGTGACAGTGTGACCTCTCCTC 1351
|||||
Db 180 GIYAspserGlyGlyProleu187

RESULT 32
US-08-615-271-12
Sequence 12, Application US/08615271
Patent No. 5981503
GENERAL INFORMATION:
APPLICANT: LOOSMORE, Sheena M.
APPLICANT: YONG, Yan-Ping
APPLICANT: CHONG, Pele
APPLICANT: OOMEN, Raymond P.
APPLICANT: KLEIN, Michael H.
TITLE OF INVENTION: ANALOG OF HAEMOPHILUS HIN47 WITH REDUCED
TITLE OF INVENTION: PROTEASE ACTIVITY
NUMBER OF SEQUENCES: 23
CORRESPONDENCE ADDRESS:
ADDRESSEE: Sim & McBurney
STREET: 6th Floor, 330 University Avenue
CITY: Toronto
STATE: Ontario
COUNTRY: Canada
ZIP: M5G 1R7
COMPUTER READABLE FORM:
MEDIUM TYPE: floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/615,271
FILING DATE: 20-JUN-1996
CLASSIFICATION: 514
ATTORNEY/AGENT INFORMATION:
NAME: Stewart, Michael I
REGISTRATION NUMBER: 24,973
```

```
REFERENCE/DOCKET NUMBER: 1038-580
TELECOMMUNICATION INFORMATION:
TELEPHONE: (416) 595-1155
TELEFAX: (416) 595-1163
INFORMATION FOR SEQ ID NO: 12:
SEQUENCE CHARACTERISTICS:
LENGTH: 224 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-615-271-12

Alignment Scores:
Pred. No.: 8.9 Length: 224
Score: 8.00 Matches: 8
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 1.68% Indels: 0
DB: 2 Gaps: 0

US-10-020-441-1 (1-1477) x US-08-615-271-12 (1-224)
QY 1328 GGTGACAGTGTGACCTCTCCTC 1351
|||||
Db 180 GIYAspserGlyGlyProleu187

RESULT 33
US-09-074-660-12
Sequence 12, Application US/09074660
Patent No. 6020183
GENERAL INFORMATION:
APPLICANT: LOOSMORE, Sheena M.
APPLICANT: YONG, Yan-Ping
APPLICANT: CHONG, Pele
APPLICANT: OOMEN, Raymond P.
APPLICANT: KLEIN, Michael H.
TITLE OF INVENTION: Analog of Haemophilus Hin47 Protein with
TITLE OF INVENTION: Reduced Protease Activity
NUMBER OF SEQUENCES: 23
CORRESPONDENCE ADDRESS:
ADDRESSEE: Sim & McBurney
STREET: 6th Floor, 330 University Avenue
CITY: Toronto
STATE: Ontario
COUNTRY: Canada
ZIP: M5G 1R7
COMPUTER READABLE FORM:
MEDIUM TYPE: floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/074,660
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/487,167
FILING DATE: 07-JUN-1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/296,149
FILING DATE: 26-AUG-1994
APPLICATION NUMBER: US 08/278,091
FILING DATE: 21-JUL-1994
ATTORNEY/AGENT INFORMATION:
NAME: Stewart, Michael I.
REGISTRATION NUMBER: 24,973
REFERENCE/DOCKET NUMBER: 1038-731 MIS:jfb
TELECOMMUNICATION INFORMATION:
TELEPHONE: (416) 595-1155
TELEFAX: (416) 595-1163
INFORMATION FOR SEQ ID NO: 12:
SEQUENCE CHARACTERISTICS:
```

LENGTH: 224 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
US-09-074-660-12

Alignment Scores:  
Pred. No.: 8.9 Length: 224  
Score: 8.00 Matches: 8  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 1.68% Indels: 0  
DB: 3 Gaps: 0

US-10-020-441-1 (1-1477) x US-09-074-660-12 (1-224)

QY 1328 GGTGACAGTGTGACCTCTCCTC 1351  
|||||  
Db 180 GlyAspSerGlyGlyProLeuLeu 187

RESULT 34  
US-09-074-659-12  
Sequence 12, Application US/09074659  
Patent No. 6025342  
GENERAL INFORMATION:  
APPLICANT: LOOSMORE, Sheena M.  
APPLICANT: YANG, Yan-Ping  
APPLICANT: CHONG, Pele  
APPLICANT: COHEN, Raymond P.  
APPLICANT: KLEIN, Michel H.  
TITLE OF INVENTION: Analog of Haemophilus Hinf47 Protein with  
NUMBER OF SEQUENCES: 23  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Sim & McBurney  
STREET: 6th Floor, 330 University Avenue  
CITY: Toronto  
STATE: Ontario  
COUNTRY: Canada  
ZIP: M5G 1R7  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/074,659  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/487,167  
FILING DATE: 07-JUN-1995  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/296,149  
FILING DATE: 26-AUG-1994  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/278,091  
FILING DATE: 21-JUL-1994  
ATTORNEY/AGENT INFORMATION:  
NAME: Stewart, Michael I.  
REGISTRATION NUMBER: 24,973  
REFERENCE/DOCKET NUMBER: 1038-730 MTS:jd  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (416) 595-1155  
TELEFAX: (416) 595-1163  
INFORMATION FOR SEQ ID NO: 12:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 224 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
US-09-074-659-12

Alignment Scores:  
Pred. No.: 8.9 Length: 224  
Score: 8.00 Matches: 8  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 1.68% Indels: 0  
DB: 3 Gaps: 0

US-10-020-441-1 (1-1477) x US-09-074-659-12 (1-224)

QY 1328 GGTGACAGTGTGACCTCTCCTC 1351  
|||||  
Db 180 GlyAspSerGlyGlyProLeuLeu 187

RESULT 35  
US-08-906-769-17  
Sequence 16, Application US/0806769  
Patent No. 6077687  
GENERAL INFORMATION:  
APPLICANT: Grieve, Robert B.  
APPLICANT: Rushlow, Keith E.  
APPLICANT: Wu Hunter, Shirley  
APPLICANT: Frank, Glenn R.  
APPLICANT: Stiegler, Gary  
APPLICANT: Gaines, Patrick J.  
APPLICANT: Silver, Gary  
TITLE OF INVENTION: FLEA PROTEASE PROTEINS, NUCLEIC ACID  
NUMBER OF SEQUENCES: 190  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Sheridan Ross & McIntosh  
STREET: 1700 Lincoln Street, Suite 3500  
CITY: Denver  
STATE: Colorado  
COUNTRY: USA  
ZIP: 80203  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/906,769  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/639,075  
FILING DATE: 24-APR-1996  
ATTORNEY/AGENT INFORMATION:  
NAME: Connell, Gary J.  
REGISTRATION NUMBER: 32,020  
REFERENCE/DOCKET NUMBER: 2618-25-C2  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (303) 863-9700  
TELEFAX: (303) 863-0223  
INFORMATION FOR SEQ ID NO: 17:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 224 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-906-769-17

Alignment Scores:  
Pred. No.: 8.9 Length: 224  
Score: 8.00 Matches: 8  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 1.68% Indels: 0  
DB: 3 Gaps: 0

US-10-020-441-1 (1-1477) x US-08-906-769-17 (1-224)

OY 1328 GGTGACAGTGTGACCTCTCCTC 1351  
|||||  
Db 177 GYASPSERGLYGLYPROLEUEN 184

RESULT 36  
US-09-106-468-12  
: Sequence 12, Application US/09106468  
: Patent No. 6114125  
: GENERAL INFORMATION:  
: APPLICANT: LOOSMORE, Sheena M.  
: APPLICANT: YANG, Yan-Ping  
: APPLICANT: CHONG, Pele  
: APPLICANT: OOMEN, Raymond P.  
: APPLICANT: KLEIN, Michel H.  
: TITLE OF INVENTION: ANALOG OF HAEMOPHILUS HIN47 WITH REDUCED  
: TITLE OF INVENTION: PROTEASE ACTIVITY  
: NUMBER OF SEQUENCES: 23  
: CORRESPONDENCE ADDRESS:  
: ADDRESSEE: Slim & McBurney  
: STREET: 6th Floor, 330 University Avenue  
: CITY: Toronto  
: STATE: Ontario  
: COUNTRY: Canada  
: ZIP: M5G 1R7  
: COMPUTER READABLE FORM:  
: MEDIUM TYPE: Floppy disk  
: COMPUTER: IBM PC compatible  
: OPERATING SYSTEM: PC-DOS/MS-DOS  
: SOFTWARE: PatentIn Release #1.0, Version #1.30  
: CURRENT APPLICATION DATA:  
: APPLICATION NUMBER: US/09/106,468  
: FILING DATE:  
: CLASSIFICATION:  
: PRIOR APPLICATION DATA:  
: APPLICATION NUMBER: US 08/615,271  
: FILING DATE: 20-JUN-1996  
: CLASSIFICATION:  
: ATTORNEY/AGENT INFORMATION:  
: NAME: Stewart, Michael I  
: REGISTRATION NUMBER: 24,973  
: REFERENCE/DOCKET NUMBER: 1038-825  
: TELECOMMUNICATION INFORMATION:  
: TELEPHONE: (416) 595-1155  
: TELEFAX: (416) 595-1163  
: INFORMATION FOR SEQ ID NO: 12:  
: SEQUENCE CHARACTERISTICS:  
: LENGTH: 224 amino acids  
: TYPE: amino acid  
: STRANDEDNESS: single  
: TOPOLOGY: linear  
: US-09-106-468-12

Alignment Scores:  
Pred. No.: 8.9 Length: 224  
Score: 8.00 Matches: 8  
Percent Similarity: 100.00% Conservative: 0  
Best local Similarity: 100.00% Mismatches: 0  
Query Match: 1.68% Indels: 0  
DB: 3 Gaps: 0

US-10-020-441-1 (1-1477) x US-09-106-468-12 (1-224)

OY 1328 GGTGACAGTGTGACCTCTCCTC 1351  
|||||  
Db 180 GYASPSERGLYGLYPROLEUEN 187

RESULT 37  
US-08-906-616-17  
: Sequence 17, Application US/08906616  
: Patent No. 6121035  
: GENERAL INFORMATION:  
: APPLICANT: Grieve, Robert B.  
: APPLICANT: Ruslow, Keith E.

: APPLICANT: Wu Hunter, Shirley  
: APPLICANT: Frank, Glenn R.  
: APPLICANT: Stiegler, Gary  
: APPLICANT: Gaines, Patrick J.  
: APPLICANT: Silver, Gary  
: TITLE OF INVENTION: FLEA AMINOPEPTIDASE PROTEINS AND USES THEREOF  
: NUMBER OF SEQUENCES: 190  
: CORRESPONDENCE ADDRESS:  
: ADDRESSEE: Sheridan Ross P.C.  
: STREET: 1700 Lincoln Street, Suite 3500  
: CITY: Denver  
: STATE: Colorado  
: COUNTRY: USA  
: ZIP: 80203  
: COMPUTER READABLE FORM:  
: MEDIUM TYPE: Floppy disk  
: COMPUTER: IBM PC compatible  
: OPERATING SYSTEM: PC-DOS/MS-DOS  
: SOFTWARE: PatentIn Release #1.0, Version #1.30  
: CURRENT APPLICATION DATA:  
: APPLICATION NUMBER: US/08/906,616  
: FILING DATE: 05-AUG-1997  
: CLASSIFICATION: 536  
: ATTORNEY/AGENT INFORMATION:  
: NAME: Connell, Gary J.  
: REGISTRATION NUMBER: 32,020  
: REFERENCE/DOCKET NUMBER: 2618-25-C2-3  
: TELECOMMUNICATION INFORMATION:  
: TELEPHONE: (303) 863-9223  
: TELEFAX: (303) 863-9700  
: INFORMATION FOR SEQ ID NO: 17:  
: SEQUENCE CHARACTERISTICS:  
: LENGTH: 224 amino acids  
: TYPE: amino acid  
: TOPOLOGY: linear  
: MOLECULE TYPE: protein  
: US-08-906-616-17

Alignment Scores:  
Pred. No.: 8.9 Length: 224  
Score: 8.00 Matches: 8  
Percent Similarity: 100.00% Conservative: 0  
Best local Similarity: 100.00% Mismatches: 0  
Query Match: 1.68% Indels: 0  
DB: 3 Gaps: 0

US-10-020-441-1 (1-1477) x US-08-906-616-17 (1-224)

OY 1328 GGTGACAGTGTGACCTCTCCTC 1351  
|||||  
Db 177 GYASPSERGLYGLYPROLEUEN 184

RESULT 38  
US-09-106-466A-12  
: Sequence 12, Application US/09106466A  
: Patent No. 6147057  
: GENERAL INFORMATION:  
: APPLICANT: LOOSMORE, Sheena M.  
: APPLICANT: YANG, Yan-Ping  
: APPLICANT: CHONG, Pele  
: APPLICANT: OOMEN, Raymond P.  
: APPLICANT: KLEIN, Michel H.  
: TITLE OF INVENTION: ANALOG OF HAEMOPHILUS HIN47 WITH REDUCED  
: TITLE OF INVENTION: PROTEASE ACTIVITY  
: NUMBER OF SEQUENCES: 23  
: CORRESPONDENCE ADDRESS:  
: ADDRESSEE: Slim & McBurney  
: STREET: 6th Floor, 330 University Avenue  
: CITY: Toronto  
: STATE: Ontario  
: COUNTRY: Canada  
: ZIP: M5G 1R7  
: COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/106,466A  
FILING DATE:  
CLASSIFICATION: 514514  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/615,271  
FILING DATE: 20-JUN-1996  
CLASSIFICATION: 514514  
ATTORNEY/AGENT INFORMATION:  
NAME: Stewart, Michael I  
REGISTRATION NUMBER: 24,973  
REFERENCE/DOCKET NUMBER: 1038-826  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (416) 595-1155  
TELEFAX: (416) 595-1163  
INFORMATION FOR SEQ ID NO: 12:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 224 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
US-09-106-466A-12

Alignment Scores:  
Pred. No.: 8.9 Length: 224  
Score: 8.00 Matches: 8  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 1.68% Indels: 0  
DB: 4 Gaps: 0

US-10-020-441-1 (1-1477) x US-09-106-466A-12 (1-224)  
QY 1328 GGTGACAGTGTGGACCTCTCCTC 1351  
DB 180 G1YAspserIyglYProLeuLeu 187

RESULT 39  
US-08-639-075A-17  
Sequence 17, Application US/08639075A  
Patent No. 6150125  
GENERAL INFORMATION:  
APPLICANT: Grieve, Robert B.  
APPLICANT: Rushlow, Keith E.  
APPLICANT: Wu Hunter, Shirley  
APPLICANT: Frank, Glenn R.  
APPLICANT: Stiegler, Gary  
APPLICANT: Gaines, Patrick J.  
APPLICANT: Silver, Gary  
TITLE OF INVENTION: FLEA PROTEASE PROTEINS, NUCLEIC ACID  
TITLE OF INVENTION: MOLECULES AND USES THEREOF  
NUMBER OF SEQUENCES: 190  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Sheridan Ross & McIntosh  
STREET: 1700 Lincoln Street, Suite 3500  
CITY: Denver  
STATE: Colorado  
COUNTRY: USA  
ZIP: 80203  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/639,075A  
FILING DATE: 24-APR-1996  
CLASSIFICATION: 424  
ATTORNEY/AGENT INFORMATION:

NAME: Connell, Gary J.  
REGISTRATION NUMBER: 32,020  
REFERENCE/DOCKET NUMBER: 2618-25-C2  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (303) 863-9700  
TELEFAX: (303) 863-0223  
INFORMATION FOR SEQ ID NO: 17:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 224 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-639-075A-17

Alignment Scores:  
Pred. No.: 8.9 Length: 224  
Score: 8.00 Matches: 8  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 1.68% Indels: 0  
DB: 4 Gaps: 0

US-10-020-441-1 (1-1477) x US-08-639-075A-17 (1-224)  
QY 1328 GGTGACAGTGTGGACCTCTCCTC 1351  
DB 177 G1YAspserIyglYProLeuLeu 184

RESULT 40  
US-09-106-467-12  
Sequence 12, Application US/09106467  
Patent No. 6153580  
GENERAL INFORMATION:  
APPLICANT: LOOSMORE, Sheena M.  
APPLICANT: YANG, Yan-Ping  
APPLICANT: CHONG, Pele  
APPLICANT: OOMEN, Raymond P.  
APPLICANT: KLEIN, Michel H.  
TITLE OF INVENTION: ANALOG OF HAEMOPHILUS H1N47 WITH REDUCED  
TITLE OF INVENTION: PROTEASE ACTIVITY  
NUMBER OF SEQUENCES: 23  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Slim & McBurney  
STREET: 6th Floor, 330 University Avenue  
CITY: Toronto  
STATE: Ontario  
COUNTRY: Canada  
ZIP: M5G 1R7  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/106,467  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/615,271  
FILING DATE: 20-JUN-1996  
CLASSIFICATION:  
ATTORNEY/AGENT INFORMATION:  
NAME: Stewart, Michael I  
REGISTRATION NUMBER: 24,973  
REFERENCE/DOCKET NUMBER: 1038-824  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (416) 595-1155  
TELEFAX: (416) 595-1163  
INFORMATION FOR SEQ ID NO: 12:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 224 amino acids  
TYPE: amino acid  
STRANDEDNESS: single



```

; TOPOLOGY: linear
US-09-106-467-12

Alignment Scores:
Pred. No.:      8.9
Score:          8.00
Percent Similarity: 100.00%
Best Local Similarity: 100.00%
Query Match:    1.68%
DB:             4

US-10-020-441-1 (1-1477) x US-09-106-467-12 (1-224)
OY  1328 GGTGACAGTGTGAGACTCTCCTC 1351
      |||
Db   180 GlyaspSerGlyGlyProLeu 187

Search completed: Apr 11 10, 2003, 07:59:17
Job time : 49.5 secs

```

**THIS PAGE BLANK (USPTO)**

















```
954 7 1.5 432 9 US-10-201-529-330 Sequence 330, App
955 7 1.5 432 9 US-10-201-530-330 Sequence 330, App
956 7 1.5 432 9 US-10-202-408-330 Sequence 330, App
957 7 1.5 432 9 US-10-202-409-330 Sequence 330, App
958 7 1.5 432 9 US-10-202-411-330 Sequence 330, App
959 7 1.5 432 9 US-10-202-472-330 Sequence 330, App
960 7 1.5 432 9 US-10-205-502-330 Sequence 330, App
961 7 1.5 432 9 US-10-205-507-330 Sequence 330, App
962 7 1.5 432 9 US-10-205-511-330 Sequence 330, App
963 7 1.5 432 9 US-10-205-902-330 Sequence 330, App
964 7 1.5 432 9 US-10-205-907-330 Sequence 330, App
965 7 1.5 432 9 US-09-888-257A-7 Sequence 7, Appli
966 7 1.5 432 9 US-10-194-456-330 Sequence 330, App
967 7 1.5 432 9 US-10-196-456-330 Sequence 330, App
968 7 1.5 432 9 US-10-198-770-330 Sequence 330, App
969 7 1.5 432 9 US-10-199-308-330 Sequence 330, App
970 7 1.5 432 9 US-10-200-617-330 Sequence 330, App
971 7 1.5 432 9 US-10-205-893-330 Sequence 330, App
972 7 1.5 432 9 US-10-205-897-330 Sequence 330, App
973 7 1.5 432 12 US-10-006-867-112 Sequence 112, App
974 7 1.5 432 12 US-10-052-586-330 Sequence 330, App
975 7 1.5 433 10 US-09-784-911-8 Sequence 8, Appli
976 7 1.5 437 10 US-09-851-588-8 Sequence 4, Appli
977 7 1.5 437 10 US-09-784-911-4 Sequence 16, Appli
978 7 1.5 444 8 US-10-099-700A-16 Sequence 2, Appli
979 7 1.5 447 10 US-09-888-615-109 Sequence 109, App
980 7 1.5 449 9 US-10-012-896-617 Sequence 617, App
981 7 1.5 449 9 US-09-895-793-617 Sequence 617, App
982 7 1.5 449 9 US-09-895-814-617 Sequence 140, App
983 7 1.5 449 9 US-10-007-280A-140 Sequence 617, App
984 7 1.5 449 10 US-09-759-143-617 Sequence 617, App
985 7 1.5 449 10 US-09-780-669-617 Sequence 617, App
986 7 1.5 449 10 US-09-822-827-617 Sequence 617, App
987 7 1.5 451 10 US-09-784-911-10 Sequence 10, Appli
988 7 1.5 453 9 US-09-978-295A-69 Sequence 69, Appli
989 7 1.5 453 9 US-09-978-697-69 Sequence 69, Appli
990 7 1.5 453 9 US-09-978-192A-69 Sequence 69, Appli
991 7 1.5 453 9 US-09-989-832A-69 Sequence 69, Appli
992 7 1.5 453 9 US-09-978-189-69 Sequence 64, Appli
993 7 1.5 453 9 US-10-174-590-64 Sequence 64, Appli
994 7 1.5 453 9 US-10-176-758-64 Sequence 64, Appli
995 7 1.5 453 9 US-10-175-737-64 Sequence 64, Appli
996 7 1.5 453 9 US-10-173-706-64 Sequence 64, Appli
997 7 1.5 453 9 US-10-175-738-64 Sequence 64, Appli
998 7 1.5 453 9 US-10-175-752-64 Sequence 64, Appli
999 7 1.5 453 9 US-10-176-482-64 Sequence 64, Appli
1000 7 1.5 453 9 US-10-176-482-64 Sequence 64, Appli
```

## ALIGNMENTS

```
RESULT 1
US-10-020-441-2
; Sequence 2, Application US/10020441
; Publication No. US20020182224A1
; GENERAL INFORMATION:
; APPLICANT: Doenhoff, Michael
; APPLICANT: Sayers, Jon
; TITLE OF INVENTION: SCHISTOSOMIASIS VACCINE
; FILE REFERENCE: Doenhoff
; CURRENT APPLICATION NUMBER: US/10/020,441
; PRIOR FILING DATE: 2001-12-18
; PRIOR APPLICATION NUMBER: US/09/413,810
; PRIOR FILING DATE: RPIOR FILING DATE: 1999-10-07
; NUMBER OF SEQ ID NOS: 3
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2
; LENGTH: 106
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: protein
US-10-020-441-2
```

```
Alignment Scores:
Pred. No.: 2.2e-97 Length: 106
Score: 106.00 Matches: 106
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 22.22% Indels: 0
DB: 9 Gaps: 0
```

US-10-020-441-1 (1-1477) x US-10-020-441-2 (1-106)

```
QY 310 GTTTCATTTCCACACCTGAGATGCGACCAAGCAGCATCCACCACCGCTTGA 369
|||||
Db 1 ValSerPheLeuThrLeuArgAsnGlyAspGlnGlnGlyIleHisGlnProSerGly 20
370 GTTAAGCGGCACGAGATACATGCGCTCTGTATGTCGACGACGACGAGAGCAATC 429
|||||
Db 21 ValLysValAlaProGlyTyrMetProSerGlyMetSerAlaArgGlnArgProIle 40
QY 430 GCACGACACTCAGTGGATTTGATTTGATTTGATTTGATTTGATTTGATTTGATTT 489
|||||
Db 41 AlaGlnThrLeuSerGlyPheAspIleAlaIleValMetLeuAlaGlnMetValAsnLeu 60
490 CAGAGTGAATCAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 549
|||||
Db 61 GlnSerGlyIleArgValIleSerLeuProGlnProSerAspIleProProGlyThr 80
QY 550 GGTGTTTCATTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTT 609
|||||
Db 81 GlyAlaPheIleValGlyTyrGlyIArgAspAspAsnAspArgAspProSerArgLysAsn 100
QY 610 GGTGATATTTGAGAAA 627
|||||
Db 101 GlyGlyIleLeuLysLys 106
```

```
RESULT 2
US-10-020-441-3
; Sequence 3, Application US/10020441
; Publication No. US20020182224A1
; GENERAL INFORMATION:
; APPLICANT: Doenhoff, Michael
; APPLICANT: Sayers, Jon
; TITLE OF INVENTION: SCHISTOSOMIASIS VACCINE
; FILE REFERENCE: Doenhoff
; CURRENT APPLICATION NUMBER: US/10/020,441
; PRIOR FILING DATE: 2001-12-18
; PRIOR APPLICATION NUMBER: US/09/413,810
; PRIOR FILING DATE: RPIOR FILING DATE: 1999-10-07
; NUMBER OF SEQ ID NOS: 3
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 3
; LENGTH: 16
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: protein
US-10-020-441-3
```

```
Alignment Scores:
Pred. No.: 3.65e-07 Length: 16
Score: 16.00 Matches: 16
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 3.35% Indels: 0
DB: 9 Gaps: 0
```

US-10-020-441-1 (1-1477) x US-10-020-441-3 (1-16)

```
QY 562 GTTGGTTATGAGAGGATGATACACCGTATCGTCGTCAGTGAAGAT 609
|||||
Db 1 ValGlyTyrGlyArgAspAspAsnAspArgAspProSerArgLysAsn 16
```

RESULT 3

US-10-001-857-135  
: Sequence 135, Application US/10001857  
: Publication No. US20020183500A1  
: GENERAL INFORMATION:  
: APPLICANT: Macina, Roberto  
: APPLICANT: Recipon, Herve  
: APPLICANT: Chen, Sel-Yu  
: APPLICANT: Sun, Yongming  
: APPLICANT: Liu, Chenghua  
: TITLE OF INVENTION: Compositions and Methods Relating to Lung Specific Genes and Prot  
: FILE REFERENCE: DEX-0273  
: CURRENT APPLICATION NUMBER: US/10/001,857  
: CURRENT FILING DATE: 2001-11-20  
: PRIOR APPLICATION NUMBER: 60/252,054  
: PRIOR FILING DATE: 2000-11-20  
: NUMBER OF SEQ ID NOS: 208  
: SOFTWARE: PatentIn version 3.1  
: SEQ ID NO 135  
: LENGTH: 94  
: TYPE: PRT  
: ORGANISM: Homo sapien  
US-10-001-857-135

Alignment Scores:  
Pred. No.: 3.13 Length: 94  
Score: 9.00 Matches: 9  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 1.89% Indels: 0  
Gaps: 0  
DB: 9

US-10-020-441-1 (1-1477) x US-10-001-857-135 (1-94)  
OY 235 TGTGCGTCGTTGTGTTGTATGC 261  
DB 56 CysAlaCysValCysValCysValCys 64

RESULT 4  
US-09-820-843A-111  
: Sequence 111, Application US/09820843A  
: Publication No. US20030039963A1  
: GENERAL INFORMATION:  
: APPLICANT: Council of Scientific and Industrial Research  
: TITLE OF INVENTION: A COMPUTATIONAL METHOD FOR THE IDENTIFICATION OF CANDIDATE PROTEI  
: FILE REFERENCE: 063915  
: CURRENT APPLICATION NUMBER: US/09/820,843A  
: CURRENT FILING DATE: 2001-03-30  
: NUMBER OF SEQ ID NOS: 118  
: SOFTWARE: PatentIn version 3.0  
: SEQ ID NO 111  
: LENGTH: 71  
: TYPE: PRT  
: ORGANISM: L. major  
: FEATURE:  
: NAME/KEY: misc\_feature  
: OTHER INFORMATION: hypothetical protein P1105.01  
: NAME/KEY: misc\_feature  
: OTHER INFORMATION: g116996498  
US-09-820-843A-111

Alignment Scores:  
Pred. No.: 32.3 Length: 71  
Score: 8.00 Matches: 8  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 1.68% Indels: 0  
Gaps: 0  
DB: 9

US-10-020-441-1 (1-1477) x US-09-820-843A-111 (1-71)  
OY 238 GCGTGCCTTGTGTTGTATGC 261  
DB 238 GCGTGCCTTGTGTTGTATGC 261

DB 39 AlaCysValCysValCysValCys 46

RESULT 5  
US-09-867-550-1010  
: Sequence 1010, Application US/09867550  
: Patent No. US20020082206A1  
: GENERAL INFORMATION:  
: APPLICANT: Leach, Martin D.  
: APPLICANT: Mephan, Fuad,  
: APPLICANT: Conley, Pamela  
: APPLICANT: Law, Debbie  
: APPLICANT: Topper, James  
: TITLE OF INVENTION: No. US20020082206A1 Polynucleotides from Atherogenic Cells a  
: FILE REFERENCE: 21402-013 (Cura-313)  
: CURRENT APPLICATION NUMBER: US/09/867,550  
: CURRENT FILING DATE: 2001-09-20  
: PRIOR APPLICATION NUMBER: USN 60/208,427  
: PRIOR FILING DATE: 2000-05-30  
: NUMBER OF SEQ ID NOS: 2125  
: SOFTWARE: FastSeq for Windows Version 4.0  
: SEQ ID NO 1010  
: LENGTH: 81  
: TYPE: PRT  
: ORGANISM: Homo sapiens  
US-09-867-550-1010

Alignment Scores:  
Pred. No.: 31.9 Length: 81  
Score: 8.00 Matches: 8  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 1.68% Indels: 0  
Gaps: 0  
DB: 10

US-10-020-441-1 (1-1477) x US-09-867-550-1010 (1-81)  
OY 691 TGTGTCGTGTTGTGTTGTGTTG 714  
DB 17 CysValCysValSerValCysVal 24

RESULT 6  
US-09-820-843A-117  
: Sequence 117, Application US/09820843A  
: Publication No. US20030039963A1  
: GENERAL INFORMATION:  
: APPLICANT: Council of Scientific and Industrial Research  
: TITLE OF INVENTION: A COMPUTATIONAL METHOD FOR THE IDENTIFICATION OF CANDIDATE PRO  
: FILE REFERENCE: 063915  
: CURRENT APPLICATION NUMBER: US/09/820,843A  
: CURRENT FILING DATE: 2001-03-30  
: NUMBER OF SEQ ID NOS: 118  
: SOFTWARE: PatentIn version 3.0  
: SEQ ID NO 117  
: LENGTH: 94  
: TYPE: PRT  
: ORGANISM: L. major  
: FEATURE:  
: NAME/KEY: misc\_feature  
: OTHER INFORMATION: hypothetical protein L2743.10  
: NAME/KEY: misc\_feature  
: OTHER INFORMATION: g116433946  
US-09-820-843A-117

Alignment Scores:  
Pred. No.: 31.4 Length: 94  
Score: 8.00 Matches: 8  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 1.68% Indels: 0  
Gaps: 0  
DB: 9

US-10-020-441-1 (1-1477) x US-09-820-843A-117 (1-94)

QY 238 GCGTGGCTTGTGTGTATGC 261  
 ||||||||||||||||  
 Db 85 AlAcYsValcYsValcYsValcYs 92

RESULT 7

US-09-953-956-14  
 ; Sequence 14, Application US/09953956  
 ; Patent No. US2002072107A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Hastings, et al.  
 ; TITLE OF INVENTION: HUMAN OSTEOCLAST-DERIVED CATHEPSIN  
 ; FILE REFERENCE: P107D2D1  
 ; CURRENT APPLICATION NUMBER: US/09/953,956  
 ; PRIOR FILING DATE: 2001-09-18  
 ; PRIOR FILING DATE: 1998-12-23  
 ; NUMBER OF SEQ ID NOS: 14  
 ; SOFTWARE: PatentIn Ver. 2.1  
 ; SEQ ID NO 14  
 ; LENGTH: 255  
 ; TYPE: PRT  
 ; ORGANISM: Homo sapiens  
 US-09-953-956-14

Alignment Scores:

Pred. No.: 28.4 Length: 255  
 Score: 8.00 Matches: 8  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 1.68% Indels: 0  
 DB: 10 Gaps: 0

US-10-020-441-1 (1-1477) x US-09-953-956-14 (1-255)

QY 1328 GGTGACAGTGTGACCTCTC 1351  
 ||||||||||||||||  
 Db 199 GlyAspSerGlyGlyProLeu 206

RESULT 8

US-10-114-464-14  
 ; Sequence 14, Application US/10114464  
 ; Patent No. US20020142448A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Hastings, et al.  
 ; TITLE OF INVENTION: HUMAN OSTEOCLAST-DERIVED CATHEPSIN  
 ; FILE REFERENCE: P107D5  
 ; CURRENT APPLICATION NUMBER: US/10/114,464  
 ; PRIOR FILING DATE: 2002-04-03  
 ; PRIOR APPLICATION NUMBER: 08/553,125  
 ; PRIOR FILING DATE: 1995-11-07  
 ; PRIOR APPLICATION NUMBER: 08/208,007  
 ; PRIOR FILING DATE: 1994-03-08  
 ; NUMBER OF SEQ ID NOS: 14  
 ; SOFTWARE: PatentIn Ver. 2.1  
 ; SEQ ID NO 14  
 ; LENGTH: 255  
 ; TYPE: PRT  
 ; ORGANISM: Homo sapiens  
 US-10-114-464-14

Alignment Scores:

Pred. No.: 28.4 Length: 255  
 Score: 8.00 Matches: 8  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 1.68% Indels: 0  
 DB: 12 Gaps: 0

US-10-020-441-1 (1-1477) x US-10-114-464-14 (1-255)

QY 1328 GGTGACAGTGTGACCTCTC 1351

Db ||||||||||||||||  
 199 GlyAspSerGlyGlyProLeu 206

RESULT 9

US-09-738-626-4716  
 ; Publication 4716, Application US/09738626  
 ; Patent No. US20020197605A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: NAKAGAWA, SATOSHI  
 ; APPLICANT: MIZOGUCHI, HIROSHI  
 ; APPLICANT: ANDO, SEIKO  
 ; APPLICANT: HAYASHI, MIKIRO  
 ; APPLICANT: OCHIAI, KEIKO  
 ; APPLICANT: YOKOI, HARUHIKO  
 ; APPLICANT: TATEISHI, NAOKO  
 ; APPLICANT: SENOH, AKIHIRO  
 ; APPLICANT: IKEDA, MASATO  
 ; APPLICANT: OZAKI, AKIO  
 ; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES  
 ; FILE REFERENCE: 249-125  
 ; CURRENT APPLICATION NUMBER: US/09/738,626  
 ; PRIOR FILING DATE: 2000-12-18  
 ; PRIOR APPLICATION NUMBER: JP 99/377484  
 ; PRIOR FILING DATE: 1999-12-16  
 ; PRIOR APPLICATION NUMBER: JP 00/159162  
 ; PRIOR FILING DATE: 2000-04-07  
 ; PRIOR APPLICATION NUMBER: JP 00/280988  
 ; PRIOR FILING DATE: 2000-08-03  
 ; NUMBER OF SEQ ID NOS: 7059  
 ; SOFTWARE: PatentIn ver. 3.0  
 ; SEQ ID NO 4716  
 ; LENGTH: 278  
 ; TYPE: PRT  
 ; ORGANISM: Corynebacterium glutamicum  
 US-09-738-626-4716

Alignment Scores:

Pred. No.: 28.2 Length: 278  
 Score: 8.00 Matches: 8  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 1.68% Indels: 0  
 DB: 9 Gaps: 0

US-10-020-441-1 (1-1477) x US-09-738-626-4716 (1-278)

QY 1325 CCAGGTGACAGTGTGACCTCTC 1348  
 ||||||||||||||||  
 Db 186 ProGlyAspSerGlyGlyProLeu 193

RESULT 10

US-09-915-181A-3  
 ; Sequence 3, Application US/09915181A  
 ; Patent No. US20020098473A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: EDWARDS, ROBERT  
 ; APPLICANT: BELLOCCHIO, ELIZABETH  
 ; APPLICANT: FREMEAUX, ROBERT  
 ; APPLICANT: REIMER, RICHARD  
 ; TITLE OF INVENTION: NOVEL GLUTAMATE TRANSPORTERS  
 ; FILE REFERENCE: 305T-932610US  
 ; CURRENT APPLICATION NUMBER: US/09/915,181A  
 ; PRIOR FILING DATE: 2002-03-26  
 ; PRIOR APPLICATION NUMBER: 07/60/220,556  
 ; PRIOR FILING DATE: 2000-07-25  
 ; NUMBER OF SEQ ID NOS: 11  
 ; SOFTWARE: PatentIn version 3.0  
 ; SEQ ID NO 3  
 ; LENGTH: 850  
 ; TYPE: PRT  
 ; ORGANISM: Homo sapiens  
 US-09-915-181A-3

```
Alignment Scores:
Pred. No.: 25.2      Length: 850
Score: 8.00          Matches: 8
Percent Similarity: 100.00%      Conservative: 0
Best Local Similarity: 100.00%    Mismatches: 0
Query Match: 1.68%              Indels: 0
DB: 10                      Gaps: 0
US-10-020-441-1 (1-1477) x US-09-915-181A-3 (1-850)

OY 235 TGTGCGTGGCTTTGTGTGTGTA 258
Db 842 CysalacysValcysValcysVal 849

RESULT 11
US-09-796-294-27
; Sequence 27, Application US/09796294
; Patent No. US20020037581A1
; GENERAL INFORMATION:
; APPLICANT: O'Brien, Timothy J.
; TITLE OF INVENTION: Extracellular Serine Protease
; FILE REFERENCE: D6020CIP3
; CURRENT APPLICATION NUMBER: US/09/796,294
; CURRENT FILING DATE: 2001-02-28
; PRIOR APPLICATION NUMBER: US 09/618,259
; PRIOR FILING DATE: 2000-07-18
; NUMBER OF SEQ ID NOS: 72
; SEQ ID NO 27
; LENGTH: 9
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: Residues 208-216 of the TADG-14 protein
US-09-796-294-27

Alignment Scores:
Pred. No.: 1.35e+07      Length: 9
Score: 7.00             Matches: 7
Percent Similarity: 100.00%      Conservative: 0
Best Local Similarity: 100.00%    Mismatches: 0
Query Match: 1.47%          Indels: 0
DB: 10                  Gaps: 0
US-10-020-441-1 (1-1477) x US-09-796-294-27 (1-9)

OY 1328 GGTGACAGTGTGACCTCTC 1348
Db 3 GlyaspserGlyGlyProleu 9

RESULT 12
US-09-918-243-84
; Sequence 84, Application US/09918243
; Patent No. US20020142317A1
; GENERAL INFORMATION:
; APPLICANT: O'Brien, Timothy J.
; APPLICANT: Cannon, Martin J.
; APPLICANT: Santini, Alessandro
; TITLE OF INVENTION: Methods for the early diagnosis of ovarian cancer
; FILE REFERENCE: D6223CIP/C/D/CIP
; CURRENT APPLICATION NUMBER: US/09/918,243
; CURRENT FILING DATE: 2001-07-30
; PRIOR APPLICATION NUMBER: US
; PRIOR FILING DATE: 2001-07-13
; NUMBER OF SEQ ID NOS: 136
; SEQ ID NO 84
; LENGTH: 9
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CHAIN
; OTHER INFORMATION: Residues 201-209 of the SCCE protein
US-09-918-243-84
```

---

```
Alignment Scores:
Pred. No.: 1.35e+07      Length: 9
Score: 7.00             Matches: 7
Percent Similarity: 100.00%      Conservative: 0
Best Local Similarity: 100.00%    Mismatches: 0
Query Match: 1.47%          Indels: 0
DB: 10                  Gaps: 0
US-10-020-441-1 (1-1477) x US-09-918-243-84 (1-9)

OY 1328 GGTGACAGTGTGACCTCTC 1348
Db 3 GlyaspserGlyGlyProleu 9

RESULT 13
US-09-905-083-84
; Sequence 84, Application US/09905083
; Patent No. US20020146708A1
; GENERAL INFORMATION:
; APPLICANT: O'Brien, Timothy J.
; TITLE OF INVENTION: Compositions and Methods for the Early Diagnosis of
; FILE REFERENCE: D6223CIP/C/DIV
; CURRENT APPLICATION NUMBER: US/09/905,083
; CURRENT FILING DATE: 2001-07-13
; PRIOR APPLICATION NUMBER: US 09/502,600
; PRIOR FILING DATE: 2000-02-11
; NUMBER OF SEQ ID NOS: 136
; SEQ ID NO 84
; LENGTH: 9
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CHAIN
; OTHER INFORMATION: Residues 201-209 of the SCCE protein
US-09-905-083-84

Alignment Scores:
Pred. No.: 1.35e+07      Length: 9
Score: 7.00             Matches: 7
Percent Similarity: 100.00%      Conservative: 0
Best Local Similarity: 100.00%    Mismatches: 0
Query Match: 1.47%          Indels: 0
DB: 10                  Gaps: 0
US-10-020-441-1 (1-1477) x US-09-905-083-84 (1-9)

OY 1328 GGTGACAGTGTGACCTCTC 1348
Db 3 GlyaspserGlyGlyProleu 9

RESULT 14
US-09-791-389-75
; Sequence 75, Application US/09791389
; Publication No. US2003003273A1
; GENERAL INFORMATION:
; APPLICANT: Herath, Herath Mudiyanselage Achula Chandrasiri
; APPLICANT: Parekh, Rajesh Bhikhu
; APPLICANT: Rohlf, Christian
; APPLICANT: Terrett, Jonathan Alexander
; APPLICANT: Tyson, Kerry Louise
; TITLE OF INVENTION: Proteins, Genes and Their Use for
; TITLE OF INVENTION: Diagnosis and Treatment of Bipolar Affective Disorder (BAD)
; FILE REFERENCE: 2543-1-001 N2
; CURRENT APPLICATION NUMBER: US/09/791,389
; CURRENT FILING DATE: 2001-02-23
; PRIOR APPLICATION NUMBER: GB 0004412.3
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: GB 0030050.9
; PRIOR FILING DATE: 2000-12-08
; PRIOR APPLICATION NUMBER: US 60/254,830
```

```
; PRIOR FILING DATE: 2000-12-12
; NUMBER OF SEQ ID NOS: 308
; SOFTWARE: FASTSEQ for Windows Version 4.0
; SEQ ID NO 75
; LENGTH: 11
; TYPE: PRT
; ORGANISM: homo sapien
US-09-791-389-75

Alignment Scores:
Pred. No.: 391      Length: 11
Score: 7.00      Matches: 7
Percent Similarity: 100.00%      Conservative: 0
Best Local Similarity: 100.00%      Mismatches: 0
Query Match: 1.47%      Indels: 0
DB: 9      Gaps: 0

US-10-020-441-1 (1-1477) x US-09-791-389-75 (1-11)

QY 1328 GGTGACAGTGTGACCTCTC 1348
      |||||
Db 1 G1yaspserglyglyProleu 7

RESULT 15
US-09-791-393-75
; Sequence 75, Application US/09791393
; Publication No. US20030032200A1
; GENERAL INFORMATION:
; APPLICANT: Herath, Herath Mudiyanselage Athula Chandrasiri
; APPLICANT: Parekh, Rajesh Bhiku
; APPLICANT: Rohlf, Christian
; TITLE OF INVENTION: Proteins, Genes and Their Use for
; TITLE OF INVENTION: Diagnosis and Treatment of Bipolar Affective Disorder (BAD)
; FILE REFERENCE: 2543-1-001 N1
; CURRENT APPLICATION NUMBER: US/09/791,393
; EARLIER FILING DATE: 2002-01-02
; EARLIER APPLICATION NUMBER: GB 0004412.3
; EARLIER FILING DATE: 2000-02-24
; EARLIER APPLICATION NUMBER: GB 0030050.9
; EARLIER FILING DATE: 2000-12-08
; EARLIER APPLICATION NUMBER: US 60/254,830
; EARLIER FILING DATE: 2000-12-12
; NUMBER OF SEQ ID NOS: 308
; SOFTWARE: FASTSEQ for Windows Version 4.0
; SEQ ID NO 75
; LENGTH: 11
; TYPE: PRT
; ORGANISM: homo sapien
US-09-791-393-75

Alignment Scores:
Pred. No.: 391      Length: 11
Score: 7.00      Matches: 7
Percent Similarity: 100.00%      Conservative: 0
Best Local Similarity: 100.00%      Mismatches: 0
Query Match: 1.47%      Indels: 0
DB: 9      Gaps: 0

US-10-020-441-1 (1-1477) x US-09-791-393-75 (1-11)

QY 1328 GGTGACAGTGTGACCTCTC 1348
      |||||
Db 1 G1yaspserglyglyProleu 7

RESULT 16
US-10-125-459-18
; Sequence 18, Application US/10125459
; Publication No. US20020192800A1
; GENERAL INFORMATION:
; APPLICANT: Ni et al.
; TITLE OF INVENTION: PT005P1
; FILE REFERENCE: Serine proteases
```

```
; CURRENT APPLICATION NUMBER: US/10/125,459
; CURRENT FILING DATE: 2002-04-19
; PRIOR APPLICATION NUMBER: US/09/597,839
; PRIOR FILING DATE: 2000-06-20
; PRIOR APPLICATION NUMBER: 60/133,239
; PRIOR FILING DATE: 1999-05-07
; PRIOR APPLICATION NUMBER: 60/135,163
; PRIOR FILING DATE: 1999-05-20
; PRIOR APPLICATION NUMBER: 60/147,005
; PRIOR FILING DATE: 1999-08-03
; PRIOR APPLICATION NUMBER: 60/152,935
; PRIOR FILING DATE: 1999-09-09
; PRIOR APPLICATION NUMBER: 60/162,979
; PRIOR FILING DATE: 1999-11-01
; NUMBER OF SEQ ID NOS: 19
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 18
; LENGTH: 12
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-125-459-18

Alignment Scores:
Pred. No.: 388      Length: 12
Score: 7.00      Matches: 7
Percent Similarity: 100.00%      Conservative: 0
Best Local Similarity: 100.00%      Mismatches: 0
Query Match: 1.47%      Indels: 0
DB: 9      Gaps: 0

US-10-020-441-1 (1-1477) x US-10-125-459-18 (1-12)

QY 1328 GGTGACAGTGTGACCTCTC 1348
      |||||
Db 5 G1yaspserglyglyProleu 11

RESULT 17
US-10-125-459-19
; Sequence 19, Application US/10125459
; Publication No. US20020192800A1
; GENERAL INFORMATION:
; APPLICANT: Ni et al.
; TITLE OF INVENTION: PT005P1
; FILE REFERENCE: Serine proteases
; CURRENT APPLICATION NUMBER: US/10/125,459
; CURRENT FILING DATE: 2002-04-19
; PRIOR APPLICATION NUMBER: US/09/597,839
; PRIOR FILING DATE: 2000-06-20
; PRIOR APPLICATION NUMBER: 60/133,239
; PRIOR FILING DATE: 1999-05-07
; PRIOR APPLICATION NUMBER: 60/135,163
; PRIOR FILING DATE: 1999-05-20
; PRIOR APPLICATION NUMBER: 60/147,005
; PRIOR FILING DATE: 1999-08-03
; PRIOR APPLICATION NUMBER: 60/152,935
; PRIOR FILING DATE: 1999-09-09
; PRIOR APPLICATION NUMBER: 60/162,979
; PRIOR FILING DATE: 1999-11-01
; NUMBER OF SEQ ID NOS: 19
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 19
; LENGTH: 12
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-125-459-19

Alignment Scores:
Pred. No.: 388      Length: 12
Score: 7.00      Matches: 7
Percent Similarity: 100.00%      Conservative: 0
Best Local Similarity: 100.00%      Mismatches: 0
Query Match: 1.47%      Indels: 0
DB: 9      Gaps: 0
```

US-10-020-441-1 (1-1477) x US-10-125-459-19 (1-12)

QY 1328 GGTGACAGTGTGGACCTCTC 1348

DB 5 GlyAspSerGlyGlyProLeu 11

RESULT 18

US-10-067-761-33

; Sequence 33, Application US/10067761  
; Publication No. US20020197701A1

; GENERAL INFORMATION:

; APPLICANT: Shi et al.

; TITLE OF INVENTION: Serine protease Polynucleotides, Polypeptides, and Antibodies

; FILE REFERENCE: PTO05P4

; CURRENT APPLICATION NUMBER: US/10/067,761

; CURRENT FILING DATE: 2002-02-08

; PRIOR APPLICATION NUMBER: 09/804,156

; PRIOR FILING DATE: 2001-03-13

; PRIOR APPLICATION NUMBER: 60/189,025

; PRIOR FILING DATE: 2000-03-14

; NUMBER OF SEQ ID NOS: 38

; SOFTWARE: PatentIn Ver. 2.0

; SEQ ID NO 33

; LENGTH: 12

; TYPE: PRT

; ORGANISM: Homo sapiens

US-10-067-761-33

Alignment Scores:

Pred. No.: 388

Score: 7.00

Percent Similarity: 100.00%

Best Local Similarity: 100.00%

Query Match: 1.47%

DB: 9

Gaps: 0

US-10-020-441-1 (1-1477) x US-10-067-761-33 (1-12)

QY 1328 GGTGACAGTGTGGACCTCTC 1348

DB 5 GlyAspSerGlyGlyProLeu 11

RESULT 19

US-10-067-761-34

; Sequence 34, Application US/10067761  
; Publication No. US20020197701A1

; GENERAL INFORMATION:

; APPLICANT: Shi et al.

; TITLE OF INVENTION: Serine protease Polynucleotides, Polypeptides, and Antibodies

; FILE REFERENCE: PTO05P4

; CURRENT APPLICATION NUMBER: US/10/067,761

; CURRENT FILING DATE: 2002-02-08

; PRIOR APPLICATION NUMBER: 09/804,156

; PRIOR FILING DATE: 2001-03-13

; PRIOR APPLICATION NUMBER: 60/189,025

; PRIOR FILING DATE: 2000-03-14

; NUMBER OF SEQ ID NOS: 38

; SOFTWARE: PatentIn Ver. 2.0

; SEQ ID NO 34

; LENGTH: 12

; TYPE: PRT

; ORGANISM: Homo sapiens

US-10-067-761-34

Alignment Scores:

Pred. No.: 388

Score: 7.00

Percent Similarity: 100.00%

Best Local Similarity: 100.00%

Query Match: 1.47%

DB: 9

Gaps: 0

US-10-020-441-1 (1-1477) x US-10-067-761-34 (1-12)

QY 1328 GGTGACAGTGTGGACCTCTC 1348

DB 5 GlyAspSerGlyGlyProLeu 11

RESULT 20

US-09-879-792-16

; Sequence 16, Application US/09879792  
; Patent No. US20020061850A1

; GENERAL INFORMATION:

; APPLICANT: Xiao, Yonghong

; APPLICANT: Gedrich, Richard

; TITLE OF INVENTION: Regulation of Human Transmembrane Serine

; FILE REFERENCE: PTO05P4

; CURRENT APPLICATION NUMBER: US/09/879,792

; CURRENT FILING DATE: 2001-06-13

; PRIOR APPLICATION NUMBER: US 60/211,224

; PRIOR FILING DATE: 2000-06-13

; PRIOR APPLICATION NUMBER: US 60/283,353

; PRIOR FILING DATE: 2001-04-13

; PRIOR APPLICATION NUMBER: US 60/283,648

; PRIOR FILING DATE: 2001-04-16

; PRIOR APPLICATION NUMBER: PCT

; PRIOR FILING DATE: 2001-06-12

; NUMBER OF SEQ ID NOS: 36

; SOFTWARE: FastSeq for Windows Version 4.0

; SEQ ID NO 16

; LENGTH: 12

; TYPE: PRT

; ORGANISM: Artificial Sequence

; FEATURE:

; OTHER INFORMATION: BLOCKS BL1253G

US-09-879-792-16

Alignment Scores:

Pred. No.: 388

Score: 7.00

Percent Similarity: 100.00%

Best Local Similarity: 100.00%

Query Match: 1.47%

DB: 10

Gaps: 0

US-10-020-441-1 (1-1477) x US-09-879-792-16 (1-12)

QY 1328 GGTGACAGTGTGGACCTCTC 1348

DB 4 GlyAspSerGlyGlyProLeu 10

RESULT 21

US-09-804-156-33

; Sequence 33, Application US/09804156  
; Patent No. US20020068320A1

; GENERAL INFORMATION:

; APPLICANT: Shi et al.

; TITLE OF INVENTION: Serine protease Polynucleotides, Polypeptides, and Antibodies

; FILE REFERENCE: PTO05P4

; CURRENT APPLICATION NUMBER: US/09/804,156

; CURRENT FILING DATE: 2001-03-13

; PRIOR APPLICATION NUMBER: 60/189,025

; PRIOR FILING DATE: 2000-03-14

; NUMBER OF SEQ ID NOS: 38

; SOFTWARE: PatentIn Ver. 2.0

; SEQ ID NO 33

; LENGTH: 12

; TYPE: PRT

; ORGANISM: Homo sapiens

US-09-804-156-33

Alignment Scores:

Pred. No.: 388

Score: 7.00

Length: 12

Matches: 7

Conservative: 0

Mismatches: 0

Indels: 0

Gaps: 0

Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 1.47% Indels: 0  
DB: 10 Gaps: 0

US-10-020-441-1 (1-1477) x US-09-804-156-33 (1-12)

QY 1328 GGTGACAGTGTGGACCTCTC 1348  
|||||  
DB 5 GlyaspSerglyglyProLeu 11

RESULT 22  
US-09-804-156-34  
; Sequence 34, Application US/09804156  
; Patent No. US20020068320A1  
; GENERAL INFORMATION:  
; APPLICANT: Shi et al.  
; TITLE OF INVENTION: Serine protease Polynucleotides, polypeptides, and Antibodies  
; FILE REFERENCE: PT005P4  
; CURRENT APPLICATION NUMBER: US/09/804,156  
; PRIOR FILING DATE: 2001-03-13  
; PRIOR APPLICATION NUMBER: 60/189,025  
; NUMBER OF SEQ ID NOS: 38  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 34  
; LENGTH: 12  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-804-156-34

Alignment Scores:  
Pred. No.: 388 Length: 12  
Score: 7.00 Matches: 7  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 1.47% Indels: 0  
DB: 10 Gaps: 0

US-10-020-441-1 (1-1477) x US-09-804-156-34 (1-12)

QY 1328 GGTGACAGTGTGGACCTCTC 1348  
|||||  
DB 5 GlyaspSerglyglyProLeu 11

RESULT 23  
US-09-946-633-18  
; Sequence 18, Application US/0946633  
; Patent No. US20020119925A1  
; GENERAL INFORMATION:  
; APPLICANT: Ni et al.  
; TITLE OF INVENTION: PT005P1  
; FILE REFERENCE: Serine proteases  
; CURRENT APPLICATION NUMBER: US/09/946,633  
; PRIOR FILING DATE: 2001-09-06  
; PRIOR APPLICATION NUMBER: 09/597,839  
; PRIOR FILING DATE: 2000-06-20  
; PRIOR APPLICATION NUMBER: 60/133,239  
; PRIOR FILING DATE: 1999-05-07  
; PRIOR APPLICATION NUMBER: 60/135,163  
; PRIOR FILING DATE: 1999-05-20  
; PRIOR APPLICATION NUMBER: 60/147,005  
; PRIOR FILING DATE: 1999-08-03  
; PRIOR APPLICATION NUMBER: 60/152,935  
; PRIOR FILING DATE: 1999-09-09  
; PRIOR APPLICATION NUMBER: 60/162,979  
; NUMBER OF SEQ ID NOS: 19  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 18  
; LENGTH: 12  
; TYPE: PRT  
; ORGANISM: Homo sapiens

US-09-946-633-18

Alignment Scores:  
Pred. No.: 388 Length: 12  
Score: 7.00 Matches: 7  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 1.47% Indels: 0  
DB: 10 Gaps: 0

US-10-020-441-1 (1-1477) x US-09-946-633-18 (1-12)

QY 1328 GGTGACAGTGTGGACCTCTC 1348  
|||||  
DB 5 GlyaspSerglyglyProLeu 11

RESULT 24  
US-09-946-633-19  
; Sequence 19, Application US/0946633  
; Patent No. US20020119925A1  
; GENERAL INFORMATION:  
; APPLICANT: Ni et al.  
; TITLE OF INVENTION: PT005P1  
; FILE REFERENCE: Serine proteases  
; CURRENT APPLICATION NUMBER: US/09/946,633  
; PRIOR FILING DATE: 2001-09-06  
; PRIOR APPLICATION NUMBER: 09/597,839  
; PRIOR FILING DATE: 2000-06-20  
; PRIOR APPLICATION NUMBER: 60/133,239  
; PRIOR FILING DATE: 1999-05-07  
; PRIOR APPLICATION NUMBER: 60/135,163  
; PRIOR FILING DATE: 1999-05-20  
; PRIOR APPLICATION NUMBER: 60/147,005  
; PRIOR FILING DATE: 1999-08-03  
; PRIOR APPLICATION NUMBER: 60/152,935  
; PRIOR FILING DATE: 1999-09-09  
; PRIOR APPLICATION NUMBER: 60/162,979  
; NUMBER OF SEQ ID NOS: 19  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 19  
; LENGTH: 12  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-946-633-19

Alignment Scores:  
Pred. No.: 388 Length: 12  
Score: 7.00 Matches: 7  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 1.47% Indels: 0  
DB: 10 Gaps: 0

US-10-020-441-1 (1-1477) x US-09-946-633-19 (1-12)

QY 1328 GGTGACAGTGTGGACCTCTC 1348  
|||||  
DB 5 GlyaspSerglyglyProLeu 11

RESULT 25  
US-09-885-441-6  
; Sequence 6, Application US/09885441  
; Patent No. US20020146407A1  
; GENERAL INFORMATION:  
; APPLICANT: Xiao, Yonghong  
; TITLE OF INVENTION: Regulation of Human Eosinophil Serine  
; FILE REFERENCE: 04974,00512  
; CURRENT APPLICATION NUMBER: US/09/885,441  
; PRIOR FILING DATE: 2001-06-21  
; PRIOR APPLICATION NUMBER: US 60/212,844  
; PRIOR FILING DATE: 2000-06-21



```

: PRIOR APPLICATION NUMBER: US 60/244,171
: PRIOR FILING DATE: 2000-10-31
: PRIOR APPLICATION NUMBER: US 60/279,766
: PRIOR FILING DATE: 2001-03-30
: PRIOR APPLICATION NUMBER: PCT/
: PRIOR FILING DATE: 2001-06-20
: NUMBER OF SEQ ID NOS: 58
: SOFTWARE: FastSeq for Windows Version 4.0
: SEQ ID NO 6
: LENGTH: 14
: TYPE: PRT
: ORGANISM: Artificial Sequence
: FEATURE:
: OTHER INFORMATION: BLOCKS domain
US-09-885-441-6

Alignment Scores:
Pred. No.: 382 Length: 14
Score: 7.00 Matches: 7
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 1.47% Indels: 0
DB: 10 Gaps: 0

US-10-020-441-1 (1-1477) x US-09-885-441-6 (1-14)
Qy 1328 GGTGACAGTGTGACCTCTC 1348
Db 6 GlyaspserGlyglyProleu 12

RESULT 26
US-09-826-290-247
: Sequence 247, Application US/09826290
: Patent No. US20020164668A1
: GENERAL INFORMATION:
: APPLICANT: Durham, L. Kathryn
: APPLICANT: Friedman, David L.
: APPLICANT: Herath, Herath Mudiyanselage Athula Chandrasiri
: APPLICANT: Kimmel, Lida H.
: APPLICANT: Parekh, Rajesh Bhikhu
: APPLICANT: Potter, David M.
: APPLICANT: Rohlf, Christian
: APPLICANT: Silber, B. Michael
: APPLICANT: Stiger, Thomas R.
: APPLICANT: Sunderland, P. Trey
: APPLICANT: Townsend, Robert Reid
: APPLICANT: White, Frost
: TITLE OF INVENTION: Nucleic Acid Molecules, Polypeptides and
: TITLE OF INVENTION: Uses Therefor, Including Diagnosis and Treatment of
: FILE REFERENCE: 2572-1-001 N2
: CURRENT APPLICATION NUMBER: US/09/826,290
: PRIOR FILING DATE: 2001-04-30
: PRIOR APPLICATION NUMBER: US 60/194,504
: PRIOR FILING DATE: 2000-04-03
: PRIOR APPLICATION NUMBER: US 60/253,647
: PRIOR FILING DATE: 2000-11-28
: NUMBER OF SEQ ID NOS: 492
: SOFTWARE: FastSeq for Windows Version 4.0
: SEQ ID NO 247
: LENGTH: 18
: TYPE: PRT
: ORGANISM: homo sapien
US-09-826-290-247

Alignment Scores:
Pred. No.: 372 Length: 18
Score: 7.00 Matches: 7
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 1.47% Indels: 0
DB: 9 Gaps: 0
```

```

US-10-020-441-1 (1-1477) x US-09-826-290-247 (1-18)
Qy 1328 GGTGACAGTGTGACCTCTC 1348
Db 5 GlyaspserGlyglyProleu 11

RESULT 27
US-09-879-792-15
: Sequence 15, Application US/09879792
: Patent No. US20020061850A1
: GENERAL INFORMATION:
: APPLICANT: Xiao, Yonghong
: APPLICANT: Gedrich, Richard
: TITLE OF INVENTION: Regulation of Human Transmembrane Serine
: TITLE OF INVENTION: Protease
: FILE REFERENCE: 02973.00035
: CURRENT APPLICATION NUMBER: US/09/879,792
: PRIOR FILING DATE: 2001-06-13
: PRIOR APPLICATION NUMBER: US 60/211,224
: PRIOR FILING DATE: 2000-06-13
: PRIOR APPLICATION NUMBER: US 60/283,353
: PRIOR FILING DATE: 2001-04-16
: PRIOR APPLICATION NUMBER: US 60/283,648
: PRIOR FILING DATE: 2001-04-16
: PRIOR APPLICATION NUMBER: PCT
: PRIOR FILING DATE: 2001-06-12
: NUMBER OF SEQ ID NOS: 36
: SOFTWARE: FastSeq for Windows Version 4.0
(Docket No. US20020061850A1 LIO-81-WO)

US-10-020-441-1 (1-1477) x US-09-879-792-15 (1-23)
Qy 1328 GGTGACAGTGTGACCTCTC 1348
Db 7 GlyaspserGlyglyProleu 13

RESULT 28
US-09-879-792-22
: Sequence 22, Application US/09879792
: Patent No. US20020061850A1
: GENERAL INFORMATION:
: APPLICANT: Xiao, Yonghong
: APPLICANT: Gedrich, Richard
: TITLE OF INVENTION: Regulation of Human Transmembrane Serine
: TITLE OF INVENTION: Protease
: FILE REFERENCE: 02973.00035
: CURRENT APPLICATION NUMBER: US/09/879,792
: PRIOR FILING DATE: 2001-06-13
: PRIOR APPLICATION NUMBER: US 60/211,224
: PRIOR FILING DATE: 2000-06-13
: PRIOR APPLICATION NUMBER: US 60/283,353
: PRIOR FILING DATE: 2001-04-13
: PRIOR APPLICATION NUMBER: US 60/283,648
: PRIOR FILING DATE: 2001-04-16
: PRIOR APPLICATION NUMBER: PCT
: PRIOR FILING DATE: 2001-06-12
: NUMBER OF SEQ ID NOS: 36
: SOFTWARE: FastSeq for Windows Version 4.0
(Docket No. US20020061850A1 LIO-81-WO)
```

```
; SEQ ID NO 22
; LENGTH: 23
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: BLOCKS BL00134B
US-09-879-792-22

Alignment Scores:
Pred. No.: 363      Length: 23
Score: 7.00        Matches: 7
Percent Similarity: 100.00%  Conservative: 0
Best Local Similarity: 100.00%  Mismatches: 0
Query Match: 1.47%  Indels: 0
DB: 10             Gaps: 0

US-10-020-441-1 (1-1477) x US-09-879-792-22 (1-23)

QY 1328 GGTGACAGTGTGACCTCTC 1348
      |||||||
      5 G1yAspserg1yglProleu 11

RESULT 29
US-09-885-441-9
; Sequence 9, Application US/09885441
; Patent No. US20020146407A1
; GENERAL INFORMATION:
; APPLICANT: Xiao, Yonghong
; TITLE OF INVENTION: Regulation of Human Eosinophil Serine
; FILE REFERENCE: 04974.00512
; CURRENT APPLICATION NUMBER: US/09/885,441
; PRIOR FILING DATE: 2001-06-21
; PRIOR APPLICATION NUMBER: US 60/212,844
; PRIOR FILING DATE: 2000-06-21
; PRIOR APPLICATION NUMBER: US 60/244,171
; PRIOR FILING DATE: 2000-10-31
; PRIOR APPLICATION NUMBER: US 60/279,766
; PRIOR FILING DATE: 2001-03-30
; PRIOR APPLICATION NUMBER: PCT/
; PRIOR FILING DATE: 2001-06-20
; NUMBER OF SEQ ID NOS: 58
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 9
; LENGTH: 24
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: BLOCKS domain
US-09-885-441-9

Alignment Scores:
Pred. No.: 362      Length: 24
Score: 7.00        Matches: 7
Percent Similarity: 100.00%  Conservative: 0
Best Local Similarity: 100.00%  Mismatches: 0
Query Match: 1.47%  Indels: 0
DB: 10             Gaps: 0

US-10-020-441-1 (1-1477) x US-09-885-441-9 (1-24)

QY 1328 GGTGACAGTGTGACCTCTC 1348
      |||||||
      5 G1yAspserg1yglProleu 11

RESULT 30
US-09-826-290-120
; Sequence 120, Application US/09826290
; Patent No. US2002016468A1
; GENERAL INFORMATION:
; APPLICANT: Durham, L. Kathryn
; APPLICANT: Friedman, David L.
; APPLICANT: Herath, Herath Mudiyanselage Athula Chandrasiri
```

```
; APPLICANT: Kimmel, Lida H.
; APPLICANT: Parekh, Rajesh Bhikhu
; APPLICANT: Potter, David M.
; APPLICANT: Rohlf, Christian
; APPLICANT: Silber, B. Michael
; APPLICANT: Stieger, Thomas R.
; APPLICANT: Sunderland, P. Trey
; APPLICANT: Townsend, Robert Reid
; APPLICANT: White, Frost
; APPLICANT: Williams, Stephen A.
; TITLE OF INVENTION: Nucleic Acid Molecules, Polypeptides and
; Uses Therefor, Including Diagnosis and Treatment of
; TITLE OF INVENTION: Alzheimer's Disease
; FILE REFERENCE: 2572-1-001 N2
; CURRENT APPLICATION NUMBER: US/09/826,290
; PRIOR FILING DATE: 2001-04-30
; PRIOR APPLICATION NUMBER: US 60/194,504
; PRIOR FILING DATE: 2000-04-03
; PRIOR APPLICATION NUMBER: US 60/253,647
; PRIOR FILING DATE: 2000-11-28
; NUMBER OF SEQ ID NOS: 492
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 120
; LENGTH: 29
; TYPE: PRT
; ORGANISM: homo sapien
US-09-826-290-120

Alignment Scores:
Pred. No.: 355      Length: 29
Score: 7.00        Matches: 7
Percent Similarity: 100.00%  Conservative: 0
Best Local Similarity: 100.00%  Mismatches: 0
Query Match: 1.47%  Indels: 0
DB: 9             Gaps: 0

US-10-020-441-1 (1-1477) x US-09-826-290-120 (1-29)

QY 1328 GGTGACAGTGTGACCTCTC 1348
      |||||||
      16 G1yAspserg1yglProleu 22

RESULT 31
US-09-820-893-111
; Sequence 111, Application US/09820893
; Patent No. US20020076705A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: 31 Human Secreted Proteins
; FILE REFERENCE: P2033P1
; CURRENT APPLICATION NUMBER: US/09/820,893
; PRIOR FILING DATE: 2001-03-30
; PRIOR APPLICATION NUMBER: 09/531,119
; PRIOR FILING DATE: 2000-03-20
; PRIOR APPLICATION NUMBER: 60/102,895
; PRIOR FILING DATE: 1998-10-02
; NUMBER OF SEQ ID NOS: 140
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 111
; LENGTH: 30
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-820-893-111

Alignment Scores:
Pred. No.: 354      Length: 30
Score: 7.00        Matches: 7
Percent Similarity: 100.00%  Conservative: 0
Best Local Similarity: 100.00%  Mismatches: 0
Query Match: 1.47%  Indels: 0
DB: 10             Gaps: 0

US-10-020-441-1 (1-1477) x US-09-820-893-111 (1-30)
```

QY 1328 GGTGACAGTGGTGCCTCTC 1348  
| | | | | | | | | | | | | | | | | | | | | |  
Db 14 GlyaspSerGlyGlyProLeu 20

RESULT 32  
US-09-885-441-3  
: Sequence 3, Application US/09885441  
: Patent No. US20020146407A1  
: GENERAL INFORMATION:  
: APPLICANT: XIAO, Yonghong  
: TITLE OF INVENTION: Regulation of Human Eosinophil Serine  
: FILE REFERENCE: 04974.00512  
: CURRENT APPLICATION NUMBER: US/09/885.441  
: PRIORITY FILING DATE: 2001-06-21, 844  
: PRIOR APPLICATION NUMBER: US 60/212,844  
: PRIOR FILING DATE: 2000-06-21  
: PRIOR APPLICATION NUMBER: US 60/244,171  
: PRIOR FILING DATE: 2000-10-31  
: PRIOR APPLICATION NUMBER: US 60/279,766  
: PRIOR FILING DATE: 2001-03-30  
: PRIOR APPLICATION NUMBER: PCT/  
: NUMBER OF SEQ ID NOS: 58  
: SOFTWARE: FastSeq for Windows Version 4.0  
: SEQ ID NO 3  
: LENGTH: 35  
: TYPE: PRT  
: ORGANISM: Artificial Sequence  
: FEATURE:  
: OTHER INFORMATION: BLOCKS domain  
US-09-885-441-3

Alignment Scores:  
Pred. No.: 348 Length: 35  
Score: 7.00 Matches: 7  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 1.47% Indels: 0  
DB: 10 Gaps: 0

US-10-020-441-1 (1-1477) x US-09-885-441-3 (1-35)

QY 1328 GGTGACAGTGGTGCCTCTC 1348  
| | | | | | | | | | | | | | | | | | | | | |  
Db 13 GlyaspSerGlyGlyProLeu 19

RESULT 33  
US-10-072-349-115  
: Sequence 115, Application US/10072349  
: Publication No. US20030054420A1  
: GENERAL INFORMATION:  
: APPLICANT: Rosen et al.  
: TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies  
: FILE REFERENCE: P110C1  
: CURRENT APPLICATION NUMBER: US/10/072.349  
: PRIORITY FILING DATE: 2002-02-11  
: PRIOR APPLICATION removed - See file Wrapper or Palm  
: NUMBER OF SEQ ID NOS: 334  
: SOFTWARE: PatentIn Ver. 3.1  
: SEQ ID NO 115  
: LENGTH: 43  
: TYPE: PRT  
: ORGANISM: Homo sapiens  
: FEATURE:  
: NAME/KEY: MISC\_FEATURE  
: LOCATION: (28)  
: OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids  
US-10-072-349-115

Alignment Scores:  
Pred. No.: 341 Length: 43  
Score: 7.00 Matches: 7  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 1.47% Indels: 0  
DB: 10 Gaps: 0

Score: 7.00 Matches: 7  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 1.47% Indels: 0  
DB: 9 Gaps: 0

US-10-020-441-1 (1-1477) x US-10-072-349-115 (1-43)

QY 106 GTCTCAGCAGGAGCAGTACTC 126  
| | | | | | | | | | | | | | | | | | | | | |  
Db 1 ValSerThrArgAlaValLeu 7

RESULT 34  
US-09-764-855-115  
: Sequence 115, Application US/09764855  
: Patent No. US20020119919A1  
: GENERAL INFORMATION:  
: APPLICANT: Rosen et al.  
: TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies  
: FILE REFERENCE: P110  
: CURRENT APPLICATION NUMBER: US/09/764.855  
: PRIORITY FILING DATE: 2001-01-17  
: PRIOR APPLICATION data removed - consult PALM or file wrapper  
: NUMBER OF SEQ ID NOS: 334  
: SOFTWARE: PatentIn Ver. 2.0  
: SEQ ID NO 115  
: LENGTH: 43  
: TYPE: PRT  
: ORGANISM: Homo sapiens  
: FEATURE:  
: NAME/KEY: SITE  
: LOCATION: (28)  
: OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids  
US-09-764-855-115

Alignment Scores:  
Pred. No.: 341 Length: 43  
Score: 7.00 Matches: 7  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 1.47% Indels: 0  
DB: 10 Gaps: 0

US-10-020-441-1 (1-1477) x US-09-764-855-115 (1-43)

QY 106 GTCTCAGCAGGAGCAGTACTC 126  
| | | | | | | | | | | | | | | | | | | | | |  
Db 1 ValSerThrArgAlaValLeu 7

RESULT 35  
US-09-888-615-117  
: Sequence 117, Application US/09888615  
: Patent No. US20020064856A1  
: GENERAL INFORMATION:  
: APPLICANT: PLOWMAN, GREGORY  
: APPLICANT: WHYTE, DAVID  
: APPLICANT: CAENEPEEL, SEAN  
: APPLICANT: CHARYDCZAK, GLEN  
: APPLICANT: MANNING, GERARD  
: APPLICANT: SUDARSANAM, SUCHA  
: TITLE OF INVENTION: NOVEL PROTEASES  
: FILE REFERENCE: 038602/1214  
: CURRENT APPLICATION NUMBER: US/09/888.615  
: PRIORITY FILING DATE: 2001-06-26  
: PRIOR APPLICATION NUMBER: 60/214,047  
: PRIOR FILING DATE: 2000-06-26  
: NUMBER OF SEQ ID NOS: 150  
: SOFTWARE: PatentIn Ver. 2.1  
: SEQ ID NO 117  
: LENGTH: 46  
: TYPE: PRT  
: ORGANISM: Homo sapiens  
US-09-888-615-117

Alignment Scores:

Pred. No.:	339	Length:	46
Score:	7.00	Matches:	7
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	1.47%	Indels:	0
DB:	10	Gaps:	0

US-10-020-441-1 (1-1477) x US-09-888-615-117 (1-46)

QY 1328 GGTGACAGTGTGACCTCTC 1348  
|||||

Db 1 GLYSPSPERGLGYPROLEU 7

RESULT 36  
US-09-832-197-7  
; Sequence 7, Application US/09832197  
; Patent No. US20020094955A1  
; GENERAL INFORMATION:  
; APPLICANT: Young et al.  
; TITLE OF INVENTION: Plasmidogen-Like Polynucleotides, Polypeptides, and Antibodies  
; FILE REFERENCE: PT019P1  
; CURRENT APPLICATION NUMBER: US/09/832,197  
; PRIOR FILING DATE: 2001-04-11  
; PRIOR APPLICATION NUMBER: PCT/US00/27253  
; PRIOR FILING DATE: 2000-10-04  
; PRIOR APPLICATION NUMBER: 60/158,044  
; PRIOR FILING DATE: 1999-10-07  
; NUMBER OF SEQ ID NOS: 8  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 7  
; LENGTH: 52  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-832-197-7

Alignment Scores:

Pred. No.:	335	Length:	52
Score:	7.00	Matches:	7
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	1.47%	Indels:	0
DB:	10	Gaps:	0

US-10-020-441-1 (1-1477) x US-09-832-197-7 (1-52)

QY 1328 GGTGACAGTGTGACCTCTC 1348  
|||||

Db 2 GLYSPSPERGLGYPROLEU 8

RESULT 37  
US-09-864-761-44996  
; Sequence 44996, Application US/09864761  
; Patent No. US20020048763A1  
; GENERAL INFORMATION:  
; APPLICANT: Penn, Sharon G.  
; APPLICANT: Rank, David R.  
; APPLICANT: Hanzel, David K.  
; APPLICANT: Chen, Wensheng  
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR  
; FILE REFERENCE: Aeomica-X-1  
; CURRENT APPLICATION NUMBER: US/09/864,761  
; PRIOR FILING DATE: 2001-05-23  
; PRIOR APPLICATION NUMBER: US 60/180,312  
; PRIOR FILING DATE: 2000-02-04  
; PRIOR APPLICATION NUMBER: US 60/207,456  
; PRIOR FILING DATE: 2000-05-26  
; PRIOR APPLICATION NUMBER: US 09/632,366  
; PRIOR FILING DATE: 2000-08-03  
; PRIOR APPLICATION NUMBER: GB 24263.6  
; PRIOR FILING DATE: 2000-10-04

PRIOR APPLICATION NUMBER: US 60/236,359  
; PRIOR FILING DATE: 2000-09-27  
; PRIOR APPLICATION NUMBER: PCT/US01/00666  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00667  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00664  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00669  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00665  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00668  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00663  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00662  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00661  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00670  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: US 60/234,687  
; PRIOR FILING DATE: 2000-09-21  
; PRIOR APPLICATION NUMBER: US 09/608,408  
; PRIOR FILING DATE: 2000-06-30  
; PRIOR APPLICATION NUMBER: US 09/774,203  
; PRIOR FILING DATE: 2001-01-29  
; NUMBER OF SEQ ID NOS: 49117  
; SOFTWARE: Annomax Sequence Listing Engine vers. 1.1  
; SEQ ID NO 44996  
; LENGTH: 53  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
; FEATURE:  
; OTHER INFORMATION: MAP TO AC013642.3  
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 0.53  
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 0.49  
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 0.52  
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 0.49  
; OTHER INFORMATION: SWISSPROT HIT: P98064, EVALUATE 7.00e-14  
; OTHER INFORMATION: EST\_HUMAN HIT: BF354474.1, EVALUATE 2.00e-17  
US-09-864-761-44996

Alignment Scores:

Pred. No.:	334	Length:	53
Score:	7.00	Matches:	7
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	1.47%	Indels:	0
DB:	10	Gaps:	0

US-10-020-441-1 (1-1477) x US-09-864-761-44996 (1-53)

QY 1328 GGTGACAGTGTGACCTCTC 1348  
|||||

Db 1 GLYSPSPERGLGYPROLEU 7

RESULT 38  
US-09-864-761-48240  
; Sequence 48240, Application US/09864761  
; Patent No. US20020048763A1  
; GENERAL INFORMATION:  
; APPLICANT: Penn, Sharon G.  
; APPLICANT: Rank, David R.  
; APPLICANT: Hanzel, David K.  
; APPLICANT: Chen, Wensheng  
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FO  
; FILE REFERENCE: Aeomica-X-1  
; CURRENT APPLICATION NUMBER: US/09/864,761  
; PRIOR FILING DATE: 2001-05-23  
; PRIOR APPLICATION NUMBER: US 60/180,312

```
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/632,366
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: GB 24263.6
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US01/00666
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00664
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00669
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00665
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00668
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00663
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00662
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00661
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00670
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: US 09/608,408
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/774,203
; PRIOR FILING DATE: 2001-01-29
; NUMBER OF SEQ ID NOS: 49117
; SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 48240
; LENGTH: 61
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AC005026.1
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 0.47
; OTHER INFORMATION: EST_HUMAN HIT: BF204930.1, EVALUATE 2.00e-07
; OTHER INFORMATION: SWISSPROT HIT: Q61602, EVALUATE 1.00e-17
US-09-864-761-48240

Alignment Scores:
Pred. No.: 329          Length: 61
Score: 7.00           Matches: 7
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 1.48%      Indels: 0
DB: 10                Gaps: 0

US-10-020-441-1 (1-1477) x US-09-864-761-48240 (1-61)
Oy 1030 CTCGTCATCTGACGACCTCC 1010
Db 13 LeuSerIleSerProLeuSer 19

RESULT 39
US-09-867-550-98
; Sequence 98, Application US/09867550
; Patent No. US20020082206A1
; GENERAL INFORMATION:
; APPLICANT: Leach, Martin D.
; APPLICANT: Mehrtaban, Fuad,
; APPLICANT: Conley, Pamela
; APPLICANT: Law, Debbie
; APPLICANT: Topper, James
; TITLE OF INVENTION: No. US20020082206A1el Polynucleotides from Atherogenic Cells and
```

```
; TITLE OF INVENTION: Thereby
; FILE REFERENCE: 21402-013 (Cura-313)
; CURRENT APPLICATION NUMBER: US/09/867,550
; CURRENT FILING DATE: 2001-09-20
; PRIOR APPLICATION NUMBER: USSN 60/208,427
; PRIOR FILING DATE: 2000-05-30
; NUMBER OF SEQ ID NOS: 2125
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 98
; LENGTH: 63
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: VARIANT
; LOCATION: (2)
; OTHER INFORMATION: wherein Xaa may be any one of Arg or His or Leu or Pro
; NAME/KEY: VARIANT
; LOCATION: (21)
; OTHER INFORMATION: wherein Xaa may be any one of Arg or Gly
US-09-867-550-98

Alignment Scores:
Pred. No.: 328          Length: 63
Score: 7.00           Matches: 7
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 1.47%      Indels: 0
DB: 10                Gaps: 0

US-10-020-441-1 (1-1477) x US-09-867-550-98 (1-63)
Oy 1349 CTCGATCCCTTCAGGTCA 1369
Db 33 LeuProSerLeuGlnGlyPro 39

RESULT 40
US-09-764-877-1050
; Sequence 1050, Application US/09764877
; Patent No. US20020147140A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PC005
; CURRENT APPLICATION NUMBER: US/09/764,877
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - refer to PALM or file wrapper
; NUMBER OF SEQ ID NOS: 4031
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 1050
; LENGTH: 65
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (63)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-09-764-877-1050

Alignment Scores:
Pred. No.: 327          Length: 65
Score: 7.00           Matches: 7
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 1.48%      Indels: 0
DB: 10                Gaps: 0

US-10-020-441-1 (1-1477) x US-09-764-877-1050 (1-65)
Oy 509 ATCACTGTGATTCACCTGTGT 489
Db 47 IleThrIleIleProLeuGlyCys 53
```

Search completed: April 10, 2003, 08:16:14  
Job time : 66 secs

---

GenCore version 5.1.4 p5 4578  
Copyright (c) 1993 - 2003 Compugen Ltd.

# OM protein - nucleic search, using frame\_plus\_p2n model

Run on: April 8, 2003, 18:07:04 ; Search time 1474.44 Seconds

(without alignments)  
1164.319 Million cell updates/sec

Title: US-10-020-441-2

Perfect score: 552  
Sequence: 1 VSFLLTNGDQGGIHQPSG.....YGRDDNDRDPKKGILKK 106

## Scoring table:

BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Delop 6.0 , Delext 7.0

Searched: 16154066 segs, 8097743376 residues

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

## Command line parameters:

-MODEL=frame+ p2n.model -DEV=xlh  
-Q=/cgn2\_1/USPTO.spool/US10020441/runat\_02042003\_093012\_20981/app\_query.fasta\_1.462  
-DB=EST -QFMT=fastcap -SUFFIX=ret -MINMATCH=0.1 -LOOPEL=0 -LOOPEXT=0  
-UNITS=bits -START=1 -END=1 -MATRIX=biosum62 -TRANS=human40.cdi -LIST=45  
-DOCALLIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=40 -MODE=LOCAL  
-OUTFMT=pico -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000  
-USRR=US10020441.CCNC 1.1 2013 @runat\_02042003\_093012\_20981 -NCPU=6 -ICPU=3  
-NO\_XLPHY -NO\_MMAP -LARGEQUERY -NEG\_SCORES=0 -WAIT -LONGLOG -DEV.TIMEOUT=120  
-WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7  
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

## Database :

EST:\*  
1: em\_estba:\*  
2: em\_esthum:\*  
3: em\_estin:\*  
4: em\_estmu:\*  
5: em\_estov:\*  
6: em\_estpl:\*  
7: em\_estro:\*  
8: em\_hlc:\*  
9: gb\_est1:\*  
10: gb\_est2:\*  
11: gb\_hlc:\*  
12: gb\_est3:\*  
13: gb\_est4:\*  
14: gb\_est5:\*  
15: em\_estfun:\*  
16: em\_estom:\*  
17: gb\_gss:\*  
18: em\_gss\_hum:\*  
19: em\_gss\_inv:\*  
20: em\_gss\_pln:\*  
21: em\_gss\_vit:\*  
22: em\_gss\_fun:\*  
23: em\_gss\_mam:\*  
24: em\_gss\_mus:\*  
25: em\_gss\_other:\*  
26: em\_gss\_pro:\*  
27: em\_gss\_yod:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	112	20.3	604	10	AW942266
2	102.5	18.6	545	13	BI681216
3	102	18.5	558	13	BI790801
4	102	18.5	1000	13	BI413250
5	101.5	18.4	525	13	BI775716
6	101	18.3	415	9	AI158460
7	100	18.1	411	12	BE862270
8	100	18.1	474	10	AW823937
9	100	18.1	620	9	AA049080
10	100	18.1	1030	14	BM919268
11	99	17.9	512	12	BF724180
12	99	17.9	639	9	AA131142
13	99	17.9	811	12	BG697540
14	99	17.9	904	14	BQ720404
15	98.5	17.8	524	14	BM876592
16	96.5	17.5	525	13	BM508802
17	96.5	17.5	599	10	AW583794
18	96	17.4	525	9	AA561197
19	95.5	17.3	542	13	BM511314
20	95.5	17.3	547	13	BM127437
21	95.5	17.3	555	14	BM894377
22	95.5	17.3	560	13	BM126887
23	95.5	17.3	563	13	BM509798
24	95.5	17.3	566	14	BQ085897
25	95.5	17.3	567	13	BM511470
26	95.5	17.3	571	13	BM512185
27	95.5	17.3	582	14	BM967226
28	95.5	17.3	576	14	BM967014
29	95.5	17.3	591	13	BM512581
30	95.5	17.3	595	12	BE969857
31	95.5	17.3	614	13	BM310730
32	95.5	17.3	620	10	AW951195
33	95.5	17.3	629	13	BM272242
34	95.5	17.3	632	14	BQ777957
35	95.5	17.3	639	13	BI711146
36	95.5	17.3	758	10	AW583424
37	95.5	17.3	828	12	BE969722
38	95.5	17.3	854	12	BE969613
39	95.5	17.3	947	12	BG679626
40	95.5	17.3	1162	14	BM926658
41	95	17.2	861	13	BI523820
42	94	17.0	379	12	BF849396
43	93.5	16.9	522	13	BM352159
44	93.5	16.9	550	13	BM129235
45	93.5	16.9	617	10	AW582979

## ALIGNMENTS

RESULT 1  
AW942266/c  
LOCUS  
DEFINITION  
ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM  
REFERENCE  
AUTHORS

AW942266 604 bp mRNA linear EST 23-APR-2001  
LD13269.3prime LD Drosophila melanogaster embryo Bluescript  
Drosophila melanogaster cDNA clone LD13269 3, mRNA sequence.  
AW942266  
AW942266.1 GI:8119966  
EST.  
fruit fly.  
Drosophila melanogaster  
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;  
Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
Ephydroidea; Drosophilidae; Drosophila.  
1 (bases 1 to 604)  
Harvey,D., Brokstein,P., Hong,L., Evans-Holm,M., Su,C., Tsang,G.,

TITLE Lewis, S. and Rubin, G.M.  
JOURNAL BDP/HHMI Drosophila EST Project  
COMMENT Unpublished (2001)  
Other ESTs: LD13269, 5prime  
Contact: Stapleton, M.  
BDGP

Lawrence Berkeley National Lab  
One Cyclotron Rd, Berkeley, CA 94720, USA  
Fax: 510 486 6798  
Email: [http://www.fruitfly.org/EST\\_estfruitfly.berkeley.edu](http://www.fruitfly.org/EST_estfruitfly.berkeley.edu)

Based upon the presence of a XhoI site followed by a run of 14 or more T residues at the beginning of the sequence, this clone was polyadenylated. The resulting Poly-T sequence has been removed. hit genomic AB02960: Drosophila melanogaster genomic scaffold 14300013385601, complete sequence.: 02/23/2001  
Plate: LD 132 row: F column: 9  
High quality sequence stop: 531.

FEATURES  
source  
1. .604  
Location/Qualifiers

/organism="Drosophila melanogaster"  
/db\_xref="taxon:7227"  
/clone\_lib="LD13269"  
/clone\_lib="LD Drosophila melanogaster embryo Bluescript"  
/sex="male and female"  
/dev\_stage="0 to 24 hours mixed stage embryonic"  
/lab\_host="SOLR"  
/note="Organ: embryo; Vector: Bluescript SK; Site 1: EcoRI ; Site 2: XhoI; Constructed using Stratagene ZAP-cDNA Synthesis kit. Oligo dt-primed and directionally cloned at EcoRI and XhoI in Bluescript SK(+/-)"  
BASE COUNT 183 a 139 c 113 g 169 t  
ORIGIN

Alignment Scores:  
Pred. No.: 0.00565 Length: 604  
Score: 112.00 Matches: 22  
Percent Similarity: 63.64% Conservative: 6  
Best Local Similarity: 50.00% Mismatches: 16  
Query Match: 20.29% Indels: 0  
DB: 10 Gaps: 0

US-10-020-441-2 (1-106) x AM942266 (1-604)

Qy 47 pheaspllealiallevalmetleualaglnmetvalaslnleuglnserglylleargval 66  
Db 593 TATGATTTCGACCTGTATATCTTAAGTCAGCCGGTACTTGGATGATCAATTATGTT 534  
Qy 67 lleserleuproglmprosesaspillepropoproglythrglyvalphei1levalgly 86  
Db 533 ATATGTTTACCTCAACAGATGATATTCGACCCGGTAAACATGTTTTCACCGGC 474  
Qy 87 TYRGLYARGASP 90  
Db 473 TGGGGAAAGAT 462

RESULT 2  
LOCUS B1681216 545 bp mRNA linear EST 17-SEP-2001  
DEFINITION B1681216 460544 MARC 1BOV Bos taurus cDNA 5', mRNA sequence.  
ACCESSION B1681216  
VERSION B1681216.1 GI:15634133  
KEYWORDS EST.  
SOURCE cow.  
ORGANISM Bos taurus

REFERENCE  
AUTHORS  
1 (bases 1 to 545)  
Smith, T.P.L., Groesse, W.M., Freking, B.A., Roberts, A.J., Stone, R.T.,  
Caas, E., Wray, J.E., White, J., Cho, J., Fahrenkrug, S.C., Bennett,  
G.L., Heaton, M.P., Laegreid, W.W., Rohrer, G.A., Chitko-McKown, C.G.,  
Petrea, G., Holt, I., Karamycheva, S., Liang, F., Quackenbush, J. and  
Keefe, J.W.

TITLE Sequence evaluation of four pooled-tissue normalized bovine cDNA  
JOURNAL libraries and construction of a gene index for cattle  
MEDLINE Genome Res. 11 (4), 626-630 (2001)  
COMMENT 21180013  
Contact: Smith TPL

USDA, ARS, US Meat Animal Research Center  
PO Box 166, Clay Center, NE 68933-0166, USA  
Tel: 402 762 4366  
Fax: 402 762 4390  
Email: [smitht@mail.marc.usda.gov](mailto:smitht@mail.marc.usda.gov)  
Single pass sequencing. Bases called and alt trimmed with phred  
v0.980904.e. Vector identified by cross\_match with the -minscore 18  
and -mismatch 12 options.

FEATURES  
source  
1. .545  
Location/Qualifiers

/organism="Bos taurus"  
/db\_xref="taxon:9913"  
/clone\_lib="MARC 1BOV"  
/issue\_type="pooled"  
/lab\_host="DH10B"  
/note="Vector: pCMV SPORT6; Site 1: NotI; Site 2: SalI;  
Library made from pooled tissue from lymph node, ovary,  
fat, hypothalamus, and pituitary."  
BASE COUNT 91 a 198 c 164 g 91 t 1 others  
ORIGIN

Alignment Scores:  
Pred. No.: 0.0629 Length: 545  
Score: 102.50 Matches: 22  
Percent Similarity: 57.35% Conservative: 17  
Best Local Similarity: 32.35% Mismatches: 20  
Query Match: 18.57% Indels: 9  
DB: 13 Gaps: 1

US-10-020-441-2 (1-106) x B1681216 (1-545)

Qy 36 GlnATGAGPProIleAlGlnThleu-----Sergly 46  
Db 317 CAGCGCTGCGCCCATGACAGATGATCCGCCAACCACTACTACAGCGTTAAGAACGG 376  
Qy 47 pheaspllealiallevalmetleualaglnmetvalaslnleuglnserglylleargval 66  
Db 377 GCGGACATGCGCTCTGAGCTGACAGAGCTTGAAATATCTCTGGACACTCCAGCG 436  
Qy 67 lleserleuproglmprosesaspillepropoproglythrglyvalphei1levalgly 86  
Db 437 GTACACCTGCCCCCGATCGGAGACCTTTCCCGGGGAGCCAGTGTGGGTGACGGGC 496  
Qy 87 TYRGLYARGASPASPASPARG 94  
Db 497 TGGGGCAACGTGACACATGAGAGG 520

RESULT 3  
LOCUS B1790801 558 bp mRNA linear EST 12-MAR-2002  
DEFINITION B1790801 id09c06.y1 Melton Normalized Mixed Mouse Pancreas 1 NI-MMS1 Mus musculus cDNA clone IMAGE:5662475 5' similar to SW:MCT6\_MOUSE  
ACCESSION B1790801  
VERSION B1790801.1 GI:15818526  
KEYWORDS EST.  
SOURCE Mus musculus  
ORGANISM house mouse.

REFERENCE  
AUTHORS  
1 (bases 1 to 558)  
Melton, D., Brown, J., Kenty, G., Permutt, A., Lee, C., Kaestner, K.,  
Lemishka, I., Scaer, M., Brestelli, J., Gradwohl, G., Clifton, S.,



TITLE	JOURNAL	COMMENT
Hallier L., Marra M., Page D., Wylie T., Martin J., Bilstein A., Schmitt A., Theising B., Ritter E., Ronko I., Bennett J., Cerdinas M., Gibbons M., McCann R., Cole R., Tsagarelisvili R., Williams T., Jackson I. and Bowers Y. Endocrine Pancreas Consortium		
Published (2000)		
Contact: Douglas Melton, Klaus H. Kaestner, & Hiroshi Inoue		

**FEATURES**  
**Source**

BASE COUNT	112 a	171 c	151 g	124 t
ORIGIN				

Alignment Scores:	
Pred. No.:	0.0739
Score:	107.100
Percent Similarity:	62.508
Best Local Similarity:	43.758
Query Match:	18.488
B:	13
Gaps:	0
length:	558
Matches:	21
conservative:	9
mismatches:	18
indels:	0

US-10-020-441-2 (1-106) x B1790801 (1-558)

46 GlyPheAspIleAlaIleValMetLeuAlaGlnMetValAsnLeuGlnSerGlyIleArg 65

Db 330 GGGGAGACGTTGCCCTGCTGGAGCTTGAGGTCCCTGTGAATGTCTCCACCCATATCCAC 389

66 ValIleSerLeuProGlnProSerAspIleProProGlyThrGlyValPheIleVal 85

Db 390 CCATATCCCTGCCCCCTGCGGAGACCTCCCCCTGGGACATCGTGTGGGTACA 449

86 GLYLYTGLYARGASPASPASPASP 93

Db 450 GGCTGGGCGACATTGATAATGAC 473

## RESULT 4

LOCUS

mRNA sequence. -

VERSION BI413250.1 GI:15174173

**SOURCE**

## REFERENCES

JOURNAL

COMMENT:

## FEATURES

BASE COUNT	ORIGIN	constructed by BENCO Soares and P. TALLIA DOMAROV.
202 a	294 c	274 g
		229 t
		1 others

Alignment Scores:	
Pred. No.:	0.15
Score:	102.00
Percent Similarity:	65.5%
Best Local Similarity:	43.75%
Query Match:	18.48%
DB:	13
Gaps:	0
Length:	10000
Matches:	21
Conservative:	9
Mismatches:	18
Indels:	0
Indels:	0

US-10-020-441-2 (1-106) x BI413250 (1-1000)

QY 46 GlypheaspIleAlaIleValMetLeuAlaGlnMetValAsnLeuGlnSerGlyIleArg 65

Db 339 GGGGCGAGCGTTCCTGCTGGAGCTTGAGGTCCCTGTGAATGTCTCCACCCATATCCAC 398

QY 66 ValIleSerLeuProGlnProSerAspIleProProGlyThrGlyValAlaPheIleVal 85

Db 399 CCCATATCCCTGCCCCCTGCTCGGAGACCTTCCCCCTGGACATCGTGCTGGTGACA 458

QY 86 GLYTYRGlgAspAspAsnAsp 93

```

Db      459  GGCCTGGCGCAGCATGATATATGAC 482
RESULT 5
LOCUS    B1775716                      525 bp  mRNA  linear  EST 25-SEP-2001
DEFINITION 468400 MARC 2BOV Bos taurus cDNA 5', mRNA sequence.
ACCESSION B1775716
VERSION   B1775716.1  GI:15776693
KEYWORDS  EST.
SOURCE    cow.
ORGANISM  Bos taurus
           Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
           Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
           Bovidae; Bovinae; Bos.
REFERENCE 1 (bases 1 to 525)
AUTHORS   Smith,T.P.L., Grose,W.M., Freking,B.A., Roberts,A.J., Stone,R.T.,
           Casas,E., Wray,D.E., White,J., Cho,J., Fahnenkrug,S.C., Bennett,
           G.L., Heaton,M.P., Laegreid,W.W., Rohrer,G.A., Chitko-McKown,C.G.,
           Petrea,G., Holt,I., Karamycheva,S., Liang,F., Quackenbush,J. and
           Keefe,J.W.
           Sequence evaluation of four pooled-tissue normalized bovine cDNA
           libraries and construction of a gene index for cattle
           Genome Res. 11 (4), 626-630 (2001)
JOURNAL   21180013
MEDLINE
COMMENT   Contact: Smith TP,
           USDA, ARS, US Meat Animal Research Center
           PO Box 166, Clay Center, NE 68933-0166, USA
           Tel: 402 762 4366
           Fax: 402 762 4390
           Email: smith@emaill.marc.usda.gov
           Single pass sequencing. Bases called and alt trimmed with phred
           v0.980904.e. Vector identified by cross_match with the -minscore 18
           and -minmatch 12 options.
PCR PRIMERS
FORWARD: AGGAAACAGCTATGACCAT
BACKWARD: GTTTTCCAGTCACGACG
Plate: 91 row: B column: 21
Seq primer: ATTTAGTGACACTATG.
           Location/Qualifiers
           source
             1..525
               /organism="Bos taurus"
               /db_xref="taxon:9913"
               /clone_lib="MARC 2BOV"
               /tissue_type="pooled"
               /lab_host="DH10B"
               /note="Vector: pCMV SPORT6; Site 1: NotI; Site 2: SalI;
               library made from pooled tissue from testis, thymus,
               semitendinosus muscle, longissimus muscle, pancreas,
               adrenal, and endometrium."
BASE COUNT  87 a 187 c 165 g 86 t
ORIGIN
US-10-020-441-2 (1-106) x B1775716 (1-525)

Alignment Scores:
Pred. No.: 0.0784 Length: 525
Score: 101.50 Matches: 22
Percent Similarity: 57.35% Conservative: 17
Best Local Similarity: 32.35% Mismatches: 20
Query Match: 18.39% Indels: 9
DB: 13 Gaps: 1

US-10-020-441-2 (1-106) x B1775716 (1-525)

Oy      36  G|N|A|G|A|G|P|I|L|A|G|I|N|H|I|L|E|U|-----S|E|R|G|I|Y 46
Db      298  CAGCTGCTCCCTCCATCGACGATCCGCCACCCCACTACTACGCGTTAAGACGGG 357
Oy      47  P|H|E|A|P|I|A|I|L|E|V|A|L|E|T|L|E|U|A|G|I|N|E|T|V|A|I|N|S|E|R|G|I|Y|L|E|A|T|G|V|A|L 66
Db      358  GCGGACATCGCCCTGTCGAGCTGACACAGCTTGTAATATCTCCCTCGGACGATCGACCGG 417
Oy      67  I|L|E|S|E|L|E|U|P|R|O|I|N|P|R|O|S|E|A|P|I|L|E|P|R|O|I|G|I|Y|T|H|G|I|Y|V|A|L|P|H|E|I|L|E|V|A|I|G|I|Y 86
           :::::|||||
           ||:::|||||

```

```

Db      418  GTCAACCTGCCCCCTGAGTCCGAGACCTTTCCCGGGGACCGACAGTGTGGGTGACGGGC 477
Oy      87  T|Y|G|I|Y|A|G|A|P|A|P|A|S|P|A|N|S|P|A|Y|G 94
Db      478  TGGGGCAACGTGACATCGAAGG 501
RESULT 6
LOCUS    A1158460                      415 bp  mRNA  linear  EST 30-SEP-1998
DEFINITION u428403.f1 Soares mammary gland NbMMG Mus musculus cDNA clone
ACCESSION A1158460
VERSION   A1158460
KEYWORDS  IMAGE:1447205 5' similar to gb:U78542 M.musculus MCP-6 mRNA for non
           functional serine protease (MOUSE);, mRNA sequence.
SOURCE    EST.
ORGANISM  house mouse.
           Mus musculus
           Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
           Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 415)
AUTHORS   Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
           Geisels,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,
           Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,
           Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and
           Waterston,R.
           The WashU-HMNI Mouse EST Project
           Unpublished (1996)
           Contact: Marra M/Mouse EST Project
           WashU-HMNI Mouse EST Project
           Washington University School of Medicine
           4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
           Tel: 314 286 1800
           Fax: 314 286 1810
           Email: mouseest@watson.wustl.edu
           This clone is available royalty-free through LNL; contact the
           IMAGE Consortium (info@image.llnl.gov) for further information.
           MGI:920521
           Seq primer: -28ml3 rev2 ET from Amersham
           High quality sequence stop: 350.
           Location/Qualifiers
           source
             1..415
               /organism="Mus musculus"
               /strain="C57BL/6J"
               /db_xref="taxon:10090"
               /clone_lib="IMAGE:1447205"
               /clone_lib="Soares_mammary_gland_NbMMG"
               /sex="male"
               /tissue_type="mammary gland"
               /dev_stage="4 weeks"
               /lab_host="DH10B"
               /note="Organ: mammary gland; Vector: pT7T3D-Pac (Pharmacia
               ) with a modified polylinker; Site 1: Not I; Site 2: Eco
               RI; 1st strand cDNA was primed with a Not I - oligo(dT)
               primer [5'
               TGTTACCAATCTGAAGTGGAGCGCGCGAANGTTTTTTTTTTTTTTTTTTTTT
               T 3']; double-stranded cDNA was ligated to Eco RI
               adaptors (Pharmacia), digested with Not I and cloned into
               the Not I and Eco RI sites of the modified pT7T3 vector.
               RNA provided by Dr. Minoru Ko, Wayne State Univ. library
               constructed and normalized by Bento Soares and W.Fatima
               Bonaldi."
BASE COUNT  84 a 128 c 106 g 97 t
ORIGIN
US-10-020-441-2 (1-106) x A1158460 (1-415)

Alignment Scores:
Pred. No.: 0.0674 Length: 415
Score: 101.00 Matches: 21
Percent Similarity: 62.50% Conservative: 9
Best Local Similarity: 43.75% Mismatches: 18
Query Match: 18.30% Indels: 0
DB: 9 Gaps: 0

```

Oy	46	GlyPheAspAlaIleValMetLeuAlaGlnMetValAsnLysGlnSerGlyIleArg	65
Db	243	GGGGCAGACGTTCCCATCTGTGAAGCTTGAAGGCCCTCGTAATGTCGCCACCATTCCAC	302
Oy	66	VallieSerLeuProGlnInProSeArSpIleProProProGlyThrGlyValPheIleVal	85
Db	303	CCCATATCCCTCCCCCTGCCTCGAGACCTTCCCCCTGGGACATCGTGCGGTGACA	362
Oy	86	GlyTYrGLYArgAspAspAsp 93	
Db	363	GGCTGGGGCGACATTGATATGAC 386	
RESULT 7			
LOCUS	BE862270		
DEFINITION	BE862270	411 bp mRNA linear EST 29-SEP-2000	
ACCESSION	UI-M-BH0-akj-d-09-0-UI.r1 NIH_BMAP_M_S1 Mus musculus cDNA clone		
VERSION	BE862270		
KEYWORDS	BE862270.1 GI:10381038		
SOURCE	EST.		
ORGANISM	house mouse.		
REFERENCE	Mus musculus		
AUTHORS	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
TITLE	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.		
JOURNAL	1 (bases 1 to 411)		
MEDLINE	Ronald M.F., Lennon,G. and Soares,M.B.		
COMMENT	Normalization and subtraction: two approaches to facilitate gene discovery Genome Res. 6 (9), 791-806 (1996) Contact: Chin, H National Institute of Mental Health 6001 Executive Blvd. Room 7N-7190, MSC 9643, Bethesda, MD 20892-9643, USA Tel.: 301 443 1706 Fax: 301 443 9890 Email: mestr@mail.nih.gov cDNA library Preparation: M.B. Soares lab Clone distribution: Researchers may obtain BMAP cDNA clones from RESEARCH GENETICS. It should be noted that Bento Soares is generating a small number of additional specialized non-redundant arrays of BMAP cDNAs whose availability will be considered under appropriate and limited collaborative arrangements Seq primer: M13 Reverse. Location/Qualifiers 1..411 /organism="Mus musculus" /strain="C57BL/6J" /db_xref="taxon:10090" /clone="UI-M-BH0-akj-d-09-0-UI" /clone_1id="NIH_BMAP_M.S1" /dev_stage="27-32 days" /lab_host="DH10B (Life Technologies)" /note="vector: pRTD-Pac (Pharmacia) with a modified polylinker; Site 1: Not I; Site 2: Eco RI; The NIH BMAP M.S1 library is a subtracted library derived from a mixture of normalized libraries from ten regions of the mouse brain (cerebellum, brain stems, olfactory bulbs, hypothalamus, cortex, amygdala, basal ganglia, pineal gland, striatum, hippocampus). The driver used for subtraction consisted of a pool of 20,000 cDNA clones obtained from non-normalized and normalized libraries of these ten regions of the mouse brain."		
BASE COUNT	62 a 117 C 124 g 87 t		1 others
ORIGIN			
Alignment Scores:			
Pred. No.:	0.087	Length:	411
Score:	100.00	Matches:	20
Percent Similarity:	62.50%	Conservative:	10
Best Local Similarity:	41.67%	Mismatches:	18
Query Match:	18.12%	Indels:	0

[illegible]



US-10-020-441-2 (1-106) x BM919268 (1-1030)

Qy 11 GlnGlnGlyIleHsiHsiGlnProSerGlyValIyValaIaProGlyIlyrmeProSer 30  
 Db 269 CGTCAGAGATCTGGCCACCTCAGAGTGCATCGGGAGAGCAGCCTTACTACACAGA 328

Qy 31 CysMetSerAlaArgGlnArgArgProIlealaglnThrIeu-----SerG1 46  
 Db 329 CCAGCTGCTGCCGTCACAGATCATCGTGCACCCAGCTTCAATCATCATCAGACTGG 388

Qy 46 yPheAspIleAlaIleValMetIeuAlaGlnMetValAsnIeuGlnSerGlyIleArgVa 66  
 Db 389 AGCGGATATCGCCCTGCTGAGAGTGCAGAGCCCTGTGAACATCTCCAGCCGGTCACAC 448

Qy 66 IlleSerIeuProGlnProSerAspIleProProProGlyIlyrGlyValIleValG1 86  
 Db 449 GGTCAATGCTGCCCTGCTGCTGCAGACCTTCCCCCGGGAGATCCGCTGCTGACTCTGG 508

Qy 86 yTyrgIlyArgAspAspAsp 93  
 Db 509 CTGGGGCGATGTGACATGAT 530

RESULT 11  
 BF724180 512 bp mRNA linear EST 05-JAN-2001  
 LOCUS Bx01h02.y2 Human Iris cDNA (Un-normalized, unamplified): Bx Homo  
 DEFINITION BF724180  
 ACCESSION BF724180.1 GI:12040089  
 VERSION EST.  
 KEYWORDS human.  
 SOURCE Homo sapiens  
 ORGANISM human.  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 1 (bases 1 to 512)  
 REFERENCE Wistow, G.J., Bernstein, S., Behal, A. and Smith, D.  
 TITLE NEIBANK: EST analysis and bioinformatics for ocular genomics  
 JOURNAL Invest. Ophthalmol. Vis. Sci. 41, (2000) In press  
 COMMENT Contact: Wistow G  
 Section on Molecular Structure and Function  
 National Eye Institute  
 6/331, NIH, Bethesda, MD 20892-2740, USA  
 Tel: 301 402 3452  
 Fax: 301 496 0078  
 Email: graeme@helix.nih.gov  
 Plate: 01 row: h column: 02  
 Seq primer: M13RPI reverse primer (ABI).  
 Location/Qualifiers  
 1..512  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone="bx01h02"  
 /clone\_lib="Human Iris cDNA (Un-normalized, unamplified): Bx"  
 /BX  
 /tissue\_type="Iris"  
 /dev\_stage="Adult"  
 /lab\_host="HEMDH10B"  
 /note="Organ: Eye; Vector: pCMVSPORT6; Post-mortem iris tissue was pooled from 10 individuals ranging in age from 4-80 years and RNA was extracted. From this pooled sample an aliquot of 60ug of total RNA yielded 2.1ug of mRNA. A directionally cloned cDNA library in the pCMVSPORT6 vector was constructed at Life Technologies, essentially following the protocols of the Superscript Plasmid System full details of which are contained in the manufacturer's instruction manual (http://www.lifetech.com/). First strand synthesis was carried out using a Not I primer-adaptor [5'-TGACTAGTCTGATCGGAGCGGCC(CT)15-3'] . Not I blunt end inserts were cloned into the Not I/EcoR V sites in the vector. EST analysis was performed on the unamplified library at the NIH Intramural Sequencing Center (NISC)."

BASE COUNT 83 a 176 c 167 g 84 t 2 others

ORIGIN

Alignment Scores:  
 Pred. No.: 0.148 Length: 512  
 Score: 99.00 Matches: 28  
 Percent Similarity: 50.00% Conservative: 16  
 Best Local Similarity: 31.82% Mismatches: 39  
 Query Match: 17.93% Indels: 5  
 DB: 12 Gaps: 1

US-10-020-441-2 (1-106) x BF724180 (1-512)

Qy 11 GlnGlnGlyIleHsiHsiGlnProSerGlyValIyValaIaProGlyIlyrmeProSer 30  
 Db 244 CGTCAGAGATCTGGCCACCTCAGAGTGCATCGGGAGAGCAGCCTTACTACACAGA 303

Qy 31 CysMetSerAlaArgGlnArgArgProIlealaglnThrIeu-----SerG1 46  
 Db 304 CCAGCTGCTGCCGTCACAGATCATCGTGCACCCAGCTTCAATCATCATCAGACTGG 363

Qy 46 yPheAspIleAlaIleValMetIeuAlaGlnMetValAsnIeuGlnSerGlyIleArgVa 66  
 Db 364 AGCGGATATCGCCCTGCTGAGAGTGCAGAGCCCTGTGAACATCTCCAGCCGGTCACAC 423

Qy 66 IlleSerIeuProGlnProSerAspIleProProProGlyIlyrGlyValIleValG1 86  
 Db 424 GGTCAATGCTGCCCTGCTGCTGCAGACCTTCCCCCGGGAGATCCGCTGCTGACTCTGG 483

Qy 86 yTyrgIlyArgAspAspAsp 93  
 Db 484 CTGGGGCGATGTGACATGAT 505

RESULT 12  
 AA131142 639 bp mRNA linear EST 14-MAY-1997  
 LOCUS z131b01.r1 Soares\_pregnant\_uterus\_NBPU Homo sapiens cDNA clone  
 DEFINITION IMAGE:503497 5', similar to gb:M30038 ALPHA-TRYPTASE PRECURSOR (HUMAN);, mRNA sequence.  
 ACCESSION AA131142  
 VERSION AA131142.1 GI:1692796  
 KEYWORDS EST.  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 1 (bases 1 to 639)  
 REFERENCE Hallier, L., Lennon, G., Becker, M., Bernaldo, M.F., Chiappelli, B.,  
 Chisoe, S., Dietrich, N., Dubuque, T., Favellio, A., Gish, W., Hawkins  
 'M., Hultman, M., Kucaba, T., Lacy, M., Le, N., Mardis, E., Moore  
 'B., Morris, M., Parsons, J., Prange, C., Riekin, L., Rohlfing, T.,  
 Schellenberg, K., Soares, M.B., Tan, F., Thierry-Mieg, J., Trevaskis, E.,  
 Underwood, K., Wohlmann, P., Waterston, R., Wilson, R. and Marra, M.  
 Generation and analysis of 280,000 human expressed sequence tags  
 Genome Res. 6 (9), 807-828 (1996)  
 97044478  
 CONTACT: Wilson RK  
 Washington University School of Medicine  
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
 Tel: 314 286 1800  
 Fax: 314 286 1810  
 Email: est@watson.wustl.edu  
 This clone is available royalty-free through LNL; contact the  
 IMAGE Consortium (info@image.lnl.gov) for further information.  
 Insert Length: 894 Std Error: 0.00  
 Seq primer: -28M13 rev2 from Amersham  
 High quality sequence stop: 349.  
 Location/Qualifiers  
 1..639  
 /organism="Homo sapiens"  
 /db\_xref="GDB:3808405"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:503497"  
 /clone\_lib="Soares\_pregnant\_uterus\_NBPU"

FEATURES  
 source





Washington University Genome Sequencing Center For information on obtaining a clone please contact: Juliana Brown (brown@fas.harvard.edu)

MGI:200673 This sequence now available from the IMAGE consortium, for clone orders contact: info@image.llnl.gov

Seq primer: -40RP from Gibco

High quality sequence stop: 437.

Location/Qualifiers

1..525

/organism="Mus musculus"

/strain="ICR"

/db\_xref="taxon:10090"

/clone="IMAGE:5943861"

/clone\_lib="Melton Normalized Mixed Mouse Pancreas 1 N1-MMSI"

/sex="Both for embryonic & newborn, male for adult and adult islet"

/dev\_stage="Embryonic day 10.5, E12.5, E16.5, newborn, adult, mixed"

/lab\_host="DH10B"

/note="Vector: pSPORT1, Site 1: Not I; Site 2: Sal I; Five libraries representing E10.5/12.5 pancreatic bud, E16.5 pancreas, newborn pancreas, adult pancreas, and adult islets of Langerhans were separately constructed using SuperScript Plasmid Library kit (Life Technologies). cDNA was made by oligo-dT priming and size-selected by column fractionation. Libraries were amplified once on solid support and plasmid DNA from each library was prepared and mixed in equal amounts. The mixed library DNA was normalized by method #4 from Bonaldo, Lemmon, and Soares

1996 Genome Research 6:791-806; 0.5 microgram single-stranded mixed library plasmid DNA was mixed with 5 micrograms PCR product representing mixed library inserts and hybridized to an EcoT of 6. Single-stranded (unhybridized) plasmids were isolated by hydroxyapatite chromatography and used to make this library."

BASE COUNT 100 a 173 c 147 g 105 t

ORIGIN

Alignment Scores:  
Pred. No.: 0.298 Length: 525  
Score: 96.50 Matches: 24  
Percent Similarity: 48.65% Conservative: 12  
Best Local Similarity: 32.43% Mismatches: 27  
Query Match: 17.48% Indels: 11  
Gaps: 13 1

US-10-020-441-2 (1-106) x BMS08802 (1-525)

QY 16 HisGlnProSerGlyValIysValAlaProGlyTYrMetProSerCysMetSerAlaArg 35

Db 8 CATCAACTCTGGGACCTCTTGTGCTCCAGCTGGAACCGCTCGTGCTGCGCTCT-- 64

QY 36 GlnArgAProIleAlaGlnThrLeuSerGlyPheAspIleAlaIleValMetLeuAla 55

Db 65 -----GGCAATGACATCGCCCTCATCAAGCTCTCA 94

QY 56 GlnMetValAsnLeuGlnSerGlyIleArgValIleSerLeuProGlnProSerAspIle 75

Db 95 CCGACGCGCCAGCTGGGAGACCGCTCCAGCTCGCTCACTCCCTCCGCTGCTGACATC 154

QY 76 ProProGlyIleThrGlyValPheIleValGlyTYrGlyArg 89

Db 155 CTTCCCAACGAGACACCTCGTACATCACCGGCTGGGCGCT 196

RESULT 17

LOCUS

AM583794 599 bp mRNA linear EST 13-MAR-2002

DEFINITION IMAGE5637357 5' similar to gb:U00306 ELASTASE IITA PRECURSOR

(HUMAN); mRNA sequence.

ACCESSION

AM583794 GI:7260768

# KEYWORDS

## SOURCE

## ORGANISM

## REFERENCE

## AUTHORS

## TITLE

## JOURNAL

## COMMENT

## FEATURES

## SOURCE

EST.  
human.  
Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
1 (bases 1 to 599)  
Wyllie,T., Meadows,A., Clifton,S., Hillier,L., Merri,M., Page,D.,  
'E., Ronko,I., Bennett,J., Birstein,A., Schmitt,A., Theising,B., Ritter  
,R., Tsagarisshvili,R., Williams,T., Jackson,Y. and Bowers,Y.  
Washu-Harvard Pancreas Est Project  
Unpublished (2000)  
Contact: Douglas Melton, Klaus H. Kaestner, & Hiroshi Inoue  
Endocrine Pancreas Consortium  
Harvard University, Howard Hughes Medical Institute  
Dept of Molecular and Cellular Biology, 7 Divinity Ave, Cambridge,  
MA 02138  
Tel: 617-495-1812  
Fax: 617-495-8557  
Email: dmelton@biohp.harvard.edu  
Libraries were constructed by Dr. Douglas Melton  
DNA sequencing by: Washington University Genome Sequencing Center  
For information on obtaining a clone please contact: Juliana Brown (brown@fas.harvard.edu)  
This sequence now available from the IMAGE consortium, for clone orders contact: info@image.llnl.gov  
Seq primer: -40RP from Gibco  
High quality sequence stop: 373.

## FEATURES

## SOURCE

1..599  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:5637357"  
/clone\_lib="Human Pancreatic Islets"  
/tissue\_type="Islets of Langerhans"  
/dev\_stage="Adult"  
/lab\_host="DH10B"  
/note="Organ: Pancreas; Vector: pSPORT1, Site 1: Not I; Site 2: Sal I; Library constructed using SuperScript Plasmid Library kit (Life Technologies). cDNA made by oligo-dT priming. Size-selected by column fractionation; average insert size 1.08 kb. Primary library, unamplified."

BASE COUNT 115 a 185 c 174 g 125 t

ORIGIN

Alignment Scores:  
Pred. No.: 0.349 Length: 599  
Score: 96.50 Matches: 25  
Percent Similarity: 48.84% Conservative: 17  
Best Local Similarity: 29.07% Mismatches: 33  
Query Match: 17.48% Indels: 11  
Gaps: 10 1

US-10-020-441-2 (1-106) x AM583794 (1-599)

QY 4 LeuThrLeuArgAsnGlyAspGlnGlnGlyIleHisHisGlnProSerGlyValIysVal 23

Db 180 CTTGCTGTGAAGAGAGGCGCCGAGACAGTGATCCCATCAACTCTGAGAGAGCTGTTGTG 239

QY 24 AlaProGlyTYrMetProSerCysMetSerAlaArgGlnArgAProIleAlaGlnThr 43

Db 240 CATCACTCTGGAACCGCTCGTGCTGCGCTCT-- 272

QY 44 LeuSerGlyPheAspIleAlaIleValMetLeuAlaGlnMetValAsnLeuGlnSerGly 63

Db 273 -----GGCAATGACATCGCCCTCATCAAGCTCTCAACGAGCGCCATCTGGAGATGCC 326

QY 64 IleArgValIleSerLeuProGlnProSerAspIleProProGlyIleThrGlyValPhe 83

Db 327 GTCCAGCTCGCTCACTCCCTCCGCTGCTGATCACTCTTCCCAACAGACCTCGCTAC 386

QY 84 IleValGlyTYrGlyArg 89



```

Db      387 ATCACCAGCTGGGCGCT 404

RESULT 18
LOCUS   AA561197
DEFINITION
  AA561197.1 Strataegene mouse Tcell 937311 Mus musculus cDNA clone
  IMAGE:973584 5', similar to gb:M57401 Mus musculus mast cell
  protease-like protein gene, complete (MOUSE);, mRNA sequence.
ACCESSION
  AA561197
VERSION
  AA561197.1 GI:2332662
KEYWORDS
  EST.
SOURCE
  house mouse.
  Mus musculus
  Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
  Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE
  1 (bases 1 to 526)
  Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
  Geisler,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,
  Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,
  Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and
  Waterston,R.
  The WashU-HMNI Mouse EST Project
  Unpublished (1996)
  Contact: Marra M/Mouse EST Project
  WashU-HMNI Mouse EST Project
  Washington University School of Medicine
  4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
  Tel: 314 286 1800
  Fax: 314 286 1810
  Email: mouseest@wustl.edu
  This clone is available royalty-free through LBL; contact the
  IMAGE Consortium (info@image.lbl.gov) for further information.
  MGI:554312
  Seq primer: -28m13 rev1 ET from Amersham
  High quality sequence stop: 443.
  Location/Qualifiers
    source
      1..526
        /organism="Mus musculus"
        /db_xref="taxon:10090"
        /clone_image="973584"
        /clone_lib="Strataegene mouse Tcell 937311"
        /rname="Tcell1"
        /rsize="526"
        /rstart="1"
        /rstop="526"
        /rstrand="+"
        /lab_host="M30 CD4+ cells"
        /dev_stage="M30 CD4+ cells"
        /note="Organ: blood; Vector: pBluescript SK-; Site 1:
        EcoRI; Site 2: XhoI; Cloned unidirectionally. Primer:
        Oligo dT. M30 CD4+ cells. Average insert size: 1.0 kb;
        Uni-ZAP XR Vector; -5' adaptor sequence: 5' GAATTCGCGACGAG
        3' -3' adaptor sequence: 5' CTCGAGTTTTTTTTTTTTTTT 3'"
BASE COUNT      145 a      114 c      135 g      132 t
ORIGIN
  Alignment Scores:
    Pred. No.:      0.341      Length:      526
    Score:          96.00      Matches:      20
    Percent Similarity: 54.90%      Conservative: 8
    Best Local Similarity: 39.22%      Mismatches: 23
    Query Match:    17.39%      Indels:      0
    DB:             9          Gaps:      0
US-10-020-441-2 (1-106) x AA561197 (1-526)

```

```

Db      449 GGGCGACTGATTGAAAAAATGATCTACGT 481

RESULT 19
LOCUS   BM511314
DEFINITION
  BM511314.1 Human insulinoma Homo sapiens cDNA clone IMAGE:5634032
  5', similar to SW:EU3A_HUMAN P09093 ELASTASE IIA PRECURSOR ;, mRNA
  sequence.
ACCESSION
  BM511314
VERSION
  BM511314.1 GI:18682457
KEYWORDS
  EST.
SOURCE
  human.
  Homo sapiens
  Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
  Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE
  1 (bases 1 to 542)
  Welter,D., Brown,J., Kenty,G., Permutt,A., Lee,C., Kaestner,K.,
  Lemishka,I., Scaer,J., Brestelli,J., Gradwohl,G., Clifton,S.,
  Hillier,L., Marra,M., Page,D., Wylie,T., Martin,J., Blisatn,A.,
  Schmitt,A., Theising,B., Ritzer,B., Ronko,I., Bennett,J., Cardenas
  ,M., Gibbons,M., McCann,R., Cole,R., Tsagarishvili,R., Williams,T.,
  Jackson,Y. and Bowers,Y.
  Endocrine Pancreas Consortium
  Unpublished (2000)
  Other ESTs: 134605.x1
  Contact: Douglas Melton, Klaus H. Kaestner, & Hiroshi Inoue
  Endocrine Pancreas Consortium
  Harvard University, Howard Hughes Medical Institute
  Dept of Molecular and Cellular Biology, 7 Divinity Ave, Cambridge,
  MA 02138
  Tel: 617-495-1812
  Fax: 617-495-8557
  Email: dmelton@hjh.harvard.edu
  Library was constructed by Dr. J. Ferrer in vivo mass-excised to
  pBluescript SK- by Dr. H. Inoue DNA sequencing by: Washington
  University Genome Sequencing Center For information on obtaining a
  clone please contact: Dr. Hiroshi Inoue (hinoue@im.wustl.edu)
  Seq primer: -40RP from GIBCO
  High quality sequence stop: 430.
  Location/Qualifiers
    source
      1..542
        /organism="Homo sapiens"
        /db_xref="taxon:9606"
        /clone_image="5634032"
        /clone_lib="Human insulinoma"
        /rname="Insulinoma"
        /rsize="542"
        /rstart="1"
        /rstop="542"
        /rstrand="+"
        /lab_host="DH10B (phage-resistant)"
        /note="Organ: pancreas; Vector: pBluescript SK-; Site 1:
        XhoI; Site 2: EcoRI; Constructed with lambda ZAPII system
        (Stratagene) by Dr. J. Ferrer. In vivo mass-excised to
        pBluescript SK- by Dr. H. Inoue following the Washington
        University protocol
        (http://genome.wustl.edu/est/lambda_protocol.shtml).
        Please contact Hiroshi Inoue, MD/PhD for further
        information on this library (Metabolism Division, Permutt
        Laboratory, Washington University School of Medicine, Box
        8127, 660 S Euclid Ave, St. Louis, MO 63110). Note: this
        is a Washington University Pancreas EST project library."
BASE COUNT      94 a      182 c      147 g      119 t
ORIGIN
  Alignment Scores:
    Pred. No.:      0.404      Length:      542
    Score:          95.50      Matches:      25
    Percent Similarity: 47.67%      Conservative: 16
    Best Local Similarity: 29.07%      Mismatches: 34
    Query Match:    17.30%      Indels:      11
    DB:             13          Gaps:      1
US-10-020-441-2 (1-106) x BM511314 (1-542)

```

Db 255 CTGCTGTGAAGAGAGGCCCCGAGCAGGTATCCCATCACTGTGAGAGCTGTTGTG 314  
 Oy 24 AAlaproglyTyrmecProserCyemeteralargGlnargArgProIlealaglInthr 43  
 Db 315 CATCACTCTGGAACCGCTCGTGTGTGCGCTCT----- 347  
 Oy 44 LeuserglyPheaspIleAlaIleValMetIleuIaGlnMetValAsnIeuGlnsergly 63  
 Db 348 -----GGCAATGACATCGCCCTCATCAAGCTCTCAGCAGAGCCGACGCTGGAGATGCC 401  
 Oy 64 IleArgValIleSerIeuProGlnProserAspIleProProProglyThrglyValPhe 83  
 Db 402 GTCCAGCTCGCCTCACTCCCTCCGCTGGTGCATCTTCCACAGACACCTGCTAC 461  
 Oy 84 IleValGlyTyrglyArg 89  
 Db 462 ATCACCAGCTGGGCGCT 479  
 RESULT 20  
 BM127437  
 LOCUS 547 bp mRNA linear EST 12-MAR-2002  
 DEFINITION i95c01.y1 Melton Normalized Human Islet 4 N4-HIS 1 Homo sapiens  
 CDNA clone IMAGE:5674748 5' similar to SW:EL3A\_HUMAN P09093  
 ELASTASE I11A PRECURSOR ; mRNA sequence.  
 BM127437  
 ACCESSION BM127437  
 VERSION BM127437.1 GI:17121989  
 KEYWORDS EST.  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 1 (bases 1 to 547)  
 Melton,D., Brown,J., Kenty,G., Permutt,A., Lee,C., Kaestner,K.,  
 Lemishka,I., Scaerle,M., Brestelli,J., Gradwohl,G., Clifton,S.,  
 Hillier,L., Marra,M., Page,D., Wylie,T., Martin,J., Bliscain,A.,  
 Schmitt,A., Theising,B., Ritzer,E., Ronko,I., Bennett,J., Cardenas  
 'M., Gibbons,M., McCann,R., Cole,R., Tsagareishvili,R., Williams,T.,  
 Jackson,Y. and Bowers,Y.  
 Endocrine Pancreas Consortium  
 Unpublished (2000)  
 TITLE  
 JOURNAL  
 COMMENT  
 CONTACT: Douglas Melton, Klaus H. Kaestner, & Hiroshi Inoue  
 Endocrine Pancreas Consortium  
 Harvard University, Howard Hughes Medical Institute  
 Dept of Molecular and Cellular Biology, 7 Divinity Ave, Cambridge,  
 MA 02138  
 Tel: 617-495-1812  
 Fax: 617-495-8557  
 Email: dmelton@biohp.harvard.edu  
 Library was constructed by Dr. Douglas Melton DNA sequencing by:  
 Washington University Genome Sequencing Center For information on  
 obtaining a clone please contact: Juliana Brown  
 (brown@fas.harvard.edu) This sequence now available from the IMAGE  
 consortium, for clone orders contact: info@image.llnl.gov  
 High quality sequence stop: 429.  
 location/Qualifiers  
 1..547  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:5674748"  
 /clone\_lib="Melton Normalized Human Islet 4 N4-HIS 1"  
 /sex="Both"  
 /tissue\_type="Islets of Langerhans"  
 /dev\_stage="Adult"  
 /lab\_host="DH10B"  
 /note="Organ: Pancreas; Vector: pSPORT1; Site\_1: Not 1;  
 Site\_2: Sal 1; Starting library constructed using  
 Superscript Plasmid library kit (Life Technologies). CDNA  
 made by oligo-dT priming. Size-selected by column  
 fractionation; average insert size 1.08 kb. Library was  
 amplified once on solid support and plasmid DNA from  
 library was prepared. The library DNA was normalized by  
 method #4 from Bonaldo, Lennon, and Soares 1996 Genome  
 Research 6:791-806; 0.5 microgram single-stranded library

plasmid DNA was mixed with 5 micrograms PCR product  
 representing library inserts and hybridized to an Ecot of  
 20. Single-stranded (unhybridized) plasmids were isolated  
 by hydroxyapatite chromatography and used to make this  
 library."

BASE COUNT 96 a 185 c 147 g 119 t  
 ORIGIN  
 Alignment Scores:  
 Pced. No.: 0.408 Length: 547  
 Score: 95.50 Matches: 25  
 Percent Similarity: 47.67% Conservative: 16  
 Best Local Similarity: 29.07% Mismatches: 34  
 Query Match: 17.30% Indels: 11  
 DB: 13 Gaps: 1  
 US-10-020-441-2 (1-106) x BM127437 (1-547)  
 Oy 4 LeuThrLeuArgAsnGlyAspGlnGlnGlyIleHisGlnProSerGlyValIleVal 23  
 Db 247 CTGCTGTGAAGAGAGGCCCCGAGCAGGTATCCCATCACTGTGAGAGCTGTTGTG 306  
 Oy 24 AAlaproglyTyrmecProserCyemeteralargGlnargArgProIlealaglInthr 43  
 Db 307 CATCACTCTGGAACCGCTCGTGTGTGCGCTCT----- 339  
 Oy 44 LeuserglyPheaspIleAlaIleValMetIleuIaGlnMetValAsnIeuGlnsergly 63  
 Db 340 -----GGCAATGACATCGCCCTCATCAAGCTCTCAGCAGAGCCGACGCTGGAGATGCC 393  
 Oy 64 IleArgValIleSerIeuProGlnProserAspIleProProProglyThrglyValPhe 83  
 Db 394 GTCCAGCTCGCCTCACTCCCTCCGCTGGTGCATCTTCCACAGACACCTGCTAC 453  
 Oy 84 IleValGlyTyrglyArg 89  
 Db 454 ATCACCAGCTGGGCGCT 471  
 RESULT 21  
 BM894377  
 LOCUS 555 bp mRNA linear EST 11-MAR-2002  
 DEFINITION i163e06.y1 Human insulinoma Homo sapiens cDNA clone IMAGE:5635667  
 5' similar to SW:EL3A\_HUMAN P09093 ELASTASE I11A PRECURSOR ; mRNA  
 sequence.  
 BM894377  
 ACCESSION BM894377  
 VERSION BM894377.1 GI:19349845  
 KEYWORDS EST.  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 1 (bases 1 to 555)  
 Melton,D., Brown,J., Kenty,G., Permutt,A., Lee,C., Kaestner,K.,  
 Lemishka,I., Scaerle,M., Brestelli,J., Gradwohl,G., Clifton,S.,  
 Hillier,L., Marra,M., Page,D., Wylie,T., Martin,J., Bliscain,A.,  
 Schmitt,A., Theising,B., Ritzer,E., Ronko,I., Bennett,J., Cardenas  
 'M., Gibbons,M., McCann,R., Cole,R., Tsagareishvili,R., Williams,T.,  
 Jackson,Y. and Bowers,Y.  
 Endocrine Pancreas Consortium  
 Unpublished (2000)  
 TITLE  
 JOURNAL  
 COMMENT  
 CONTACT: Douglas Melton, Klaus H. Kaestner, & Hiroshi Inoue  
 Endocrine Pancreas Consortium  
 Harvard University, Howard Hughes Medical Institute  
 Dept of Molecular and Cellular Biology, 7 Divinity Ave, Cambridge,  
 MA 02138  
 Tel: 617-495-1812  
 Fax: 617-495-8557  
 Email: dmelton@biohp.harvard.edu  
 Library was constructed by Dr. J. Ferrer In vivo mass-excised to  
 pBluescript SK- by Dr. H. Inoue DNA sequencing by: Washington  
 University Genome Sequencing Center For information on obtaining a  
 clone please contact: Dr. Hiroshi Inoue (hinoue@im.wustl.edu)

Seq primer: -40RP from Gibco  
High quality sequence stop: 429.  
Location/Qualifiers  
1. 555

FEATURES  
SOURCE  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:5635667"  
/clone\_lib="Human Insulinoma"  
/issue\_type="Insulinoma"  
/lab\_host="DH10B (phage-resistant)"  
/note="Organ: pancreas; Vector: p Bluescript SK-; Site: 1; XhoI; Site 2: EcoRI; Constructed with Lambda ZapII system (Stratagene) by Dr. J. Ferrer, in vivo mass-excised to Bluescript SK- by Dr. H. Inoue following the Washington University protocol  
(http://genome.wustl.edu/est/lambda\_protocol.shtml).  
Please contact Hiroshi Inoue, MD/PhD for further information on this library (Metabolism Division, Permutt Laboratory, Washington University School of Medicine, Box 8127, 660 S Euclid Ave, St. Louis, MO 63110). Note: this is a Washington University Pancreas EST project library."

BASE COUNT 97 a 187 c 150 g 121 t  
ORIGIN

Alignment Scores:

Pred. No.: 0.416 Length: 555  
Score: 95.50 Matches: 25  
Percent Similarity: 47.67% Conservative: 16  
Best Local Similarity: 29.07% Mismatches: 34  
Query Match: 17.30% Indels: 11  
Gaps: 1

US-10-020-441-2 (1-106) x BM894377 (1-555)  
Qy 4 LeuthrleuAgaenglyAbspGingInglylEhshInProserglyVallysVal 23  
Db 256 CTTGCTGTGAAGAGAGGCCCGAGAGGTATGCCATCACTGAGAGAGCTTTTGTG 315  
Qy 24 AlaproglyTyMetProSerCyMetSerAlaArgGlnArgArgProIleAglInthr 43  
Db 316 CATCCACTGTGAACCGCTCGTGTGGCTGT----- 348  
Qy 44 LeuserglyPheaspIleAlaIleValMetleuAglMetValAsnleuGlnsergly 63  
Db 349 -----GGCATTGACATCGCCCTCATCAAGCTCTCAAGAGGCCCGAGAGTGGAGATGCC 402  
Qy 64 IleArgValIleSerleuProGlnProSeraspIleProProProglyThrglyValphe 83  
Db 403 GTCCAGCTCGCTCACTCCCTCCGCTGTGATCATCTCCCAAGACACCCGTGTAC 462  
Qy 84 IleValGlyTyArg 89  
Db 463 ATCACCGGCTGGGCGCT 480

RESULT 22  
BM126887 560 bp mRNA linear EST 12-MAR-2002  
LOCUS  
DEFINITION  
1e99f11.y1 Melton Normalized Human Islet 4 N4-HIS 1 Homo sapiens  
ELASTASE IIIA PRECURSOR ; mRNA sequence.  
ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM  
REFERENCE  
AUTHORS

BM126887  
1 (bases 1 to 560)  
Melton, D., Brown, J., Kenty, G., Permutt, A., Lee, C., Kaestner, K., Lentschka, I., Scearn, M., Brestelli, D., Gradwohl, G., Clifton, S., Hillier, L., Maira, M., Pape, D., Wylie, T., Martin, J., Blistrain, A., Schmitt, A., Theising, B., Ritzer, E., Ronko, I., Bennett, J., Cardenas, M., Gibbons, M., McCann, R., Cole, R., Tsagarelis, R., Williams, T.

TITLE  
JOURNAL  
COMMENT  
' Jackson, Y. and Bowers, Y.  
Endocrine Pancreas Consortium  
Unpublished (2000)  
Other ESTs: 1e99f11.x1  
Contact: Douglas Melton, Klaus H. Kaestner, & Hiroshi Inoue  
Endocrine Pancreas Consortium  
Harvard University, Howard Hughes Medical Institute  
Dept of Molecular and Cellular Biology, 7 Divinity Ave, Cambridge, MA 02138  
Tel: 617-495-1812  
Fax: 617-495-8557  
Email: dmelton@bioh.harvard.edu  
Library was constructed by Dr. Douglas Melton DNA sequencing by:  
Washington University Genome Sequencing Center For information on  
obtaining a clone please contact: Juliana Brown  
(brown@fas.harvard.edu) This sequence now available from the IMAGE  
consortium, for clone orders contact: info@image.llnl.gov  
Seq primer: -40RP from Gibco  
High quality sequence stop: 434.  
Location/Qualifiers  
1. 560

FEATURES  
SOURCE

/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:5675276"  
/clone\_lib="Melton Normalized Human Islet 4 N4-HIS 1"  
/sex="Both"  
/issue\_type="Islets of Langerhans"  
/dev\_stage="Adult"  
/lab\_host="DH10B"  
/note="Organ: Pancreas; Vector: pSPORT1; Site: 1; Not 1; Site 2: Sal 1; Starting library kit (Life Technologies). cDNA made by oligo-dT priming. Size-selected by column fractionation; average insert size 1.08 kb. Library was amplified once on solid support and plasmid DNA from library was prepared. The library DNA was normalized by method #4 from Bonaldo, Lennon, and Soares 1996 Genome Research 6:791-806; 0.5 microgram single-stranded library plasmid DNA was mixed with 5 micrograms PCR product representing library inserts and hybridized to an EcoC of 20. Single-stranded (unhybridized) plasmids were isolated by hydroxyapatite chromatography and used to make this library."

BASE COUNT 97 a 186 c 155 g 122 t  
ORIGIN

Alignment Scores:  
Pred. No.: 0.42 Length: 560  
Score: 95.50 Matches: 25  
Percent Similarity: 47.67% Conservative: 16  
Best Local Similarity: 29.07% Mismatches: 34  
Query Match: 17.30% Indels: 11  
Gaps: 1

US-10-020-441-2 (1-106) x BM126887 (1-560)

Qy 4 LeuthrleuAgaenglyAbspGingInglylEhshInProserglyVallysVal 23  
Db 246 CTTGCTGTGAAGAGAGGCCCGAGAGGTATGCCATCACTGAGAGAGCTTTTGTG 305  
Qy 24 AlaproglyTyMetProSerCyMetSerAlaArgGlnArgArgProIleAglInthr 43  
Db 306 CATCCACTGTGAACCGCTCGTGTGGCTGT----- 338  
Qy 44 LeuserglyPheaspIleAlaIleValMetleuAglMetValAsnleuGlnsergly 63  
Db 339 -----GGCATTGACATCGCCCTCATCAAGCTCTCAAGAGGCCCGAGAGTGGAGATGCC 392  
Qy 64 IleArgValIleSerleuProGlnProSeraspIleProProProglyThrglyValphe 83  
Db 393 GTCCAGCTCGCTCACTCCCTCCGCTGTGATCATCTCCCAAGACACCCGTGTAC 452  
Qy 84 IleValGlyTyArg 89

```

Db      453 ATCACCGGCTGGGCGCGT 470
RESULT 23
LOCUS   BM509798                      563 bp      mRNA      linear      EST 15-FEB-2002
DEFINITION 194b05.y1 HR85 islet Homo sapiens cDNA 5' similar to SW:EL3A_HUMAN
P09093 ELASTASE I11A PRECURSOR ;, mRNA sequence.
ACCESSION BM509798
VERSION   BM509798.1  GI:18680941
KEYWORDS EST.
SOURCE   human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE
AUTHORS  Melton,D., Brown,J., Kenty,G., Permutt,A., Lee,C., Kaestner,K.,
Lemishka,I., Searce,M., Brestelli,J., Gradwohl,G., Clifton,S.,
Hillier,L., Maira,M., Pape,D., Wylie,T., Martin,J., Blistain,A.,
Schmitt,A., Theising,B., Ritzer,E., Ronko,I., Bennett,J., Cardenas
,M., Gibbons,M., McCann,R., Cole,R., Tsagareishvili,R., Williams,T.,
Jackson,Y. and Bowers,Y.
TITLE     Endocrine Pancreas Consortium
JOURNAL   Unpublished (2000)
COMMENT   Other ESTs: 194b05.x1
Contact: Douglas Melton, Klaus H. Kaestner, & Hiroshi Inoue
Endocrine Pancreas Consortium
Harvard University, Howard Hughes Medical Institute
Dept of Molecular and Cellular Biology, 7 Divinity Ave, Cambridge,
MA 02138
Tel: 617-495-1812
Fax: 617-495-8557
Email: dmelton@bioph.harvard.edu
Library was constructed by Dr. Hiroshi Inoue DNA sequencing by:
Washington University Genome Sequencing Center For information on
obtaining a clone please contact: Dr. Hiroshi Inoue
(hinoue@im.wustl.edu)
Seq primer: -40RP from Gibco
High quality sequence stop: 433.
FEATURES
SOURCE    location/Qualifiers
1..563
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="HR85 islet"
/tissue_type="Purified pancreatic islet"
/lab_host="DH10B"
/note="Organ: Pancreas; Vector: pBluescript SK(-); Site_1:
NotI; Site_2: XhoI; cDNA made by oligo-dT priming.
Size-selected on agarose gel. Average insert size ~1kb. 5'
XhoI site was destroyed after directional cloning.
Amplified once. Contact information: Hiroshi Inoue, MD,
Metabolism Div. (Alan Permutt lab), Washington University
School of Medicine, Box 8127, 660 South Euclid Ave., St.
Louis, MO 63110, E-mail: hinoue@imgate.wustl.edu, Tel:
314-362-1916, Fax: 314-747-2692."
BASE COUNT 110 a 175 c 162 g 116 t
ORIGIN
Alignment Scores:
Pred. No.: 0.423 Length: 563
Score: 95.50 Matches: 25
Percent Similarity: 47.67% Conservative: 16
Best Local Similarity: 29.07% Mismatches: 34
Query Match: 17.30% Indels: 11
DB: 13 Gaps: 1
US-10-020-441-2 (1-106) x BM509798 (1-563)
Qy      4 LeuThrLeuArgAsnGlyAspGlnGlyIleHisGlnProSerGlyValVal 23
Db      138 CTTCGCTGTAAGAGGAGGCCCCCGAGGTGATCCCATCACTTCATGATGAGCTTTGTG 197
Qy      24 AlaProGlyTyrMetProSerCysMetSerAlaArgGlnArgArgProIleAlaGlnThr 43

```

```

Db      198 CATCACTCTCGAACCGCTCGTGTGTGCGCTGT----- 230
Qy      44 LeuSerGlyPheAspIleAlaIleValMetLeuAlaGlnMetValAsnLeuGlnSerGly 63
Db      231 -----GGCAATGACATCGCCCTCATCAAGCTTCACCCAGCCCGCACTGGAGATGCC 284
Qy      64 ILeaValIleSerLeuProGlnProSerAspIleProProGlyThrGlyValPhe 83
Db      285 GTCAAGCTCGGCTACTCTCCCTCCGCTGTGATACATCTTCCCAACAGACACCTGCTAC 344
Qy      84 ILeValGlyTyrGlyArg 89
Db      345 ATCACCGGCTGGGCGCGT 362
RESULT 24
LOCUS   B0085897                      566 bp      mRNA      linear      EST 05-APR-2002
DEFINITION 111e01.y1 Melton Amplified Mouse E16 5 Pancreas 3 M16S1 A Mus
musculus cDNA clone IMAGE:5941344 5' similar to SW:EL3A_HUMAN
P09093 ELASTASE I11A PRECURSOR ;, mRNA sequence.
ACCESSION B0085897
VERSION   B0085897.1  GI:20045101
KEYWORDS EST.
SOURCE   Mus musculus
ORGANISM house mouse.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciuromachi; Muridae; Murinae; Mus.
REFERENCE
AUTHORS  Melton,D., Brown,J., Kenty,G., Permutt,A., Lee,C., Kaestner,K.,
Lemishka,I., Searce,M., Brestelli,J., Gradwohl,G., Clifton,S.,
Hillier,L., Maira,M., Pape,D., Wylie,T., Martin,J., Blistain,A.,
Schmitt,A., Theising,B., Ritzer,E., Ronko,I., Bennett,J., Cardenas
,M., Gibbons,M., McCann,R., Cole,R., Tsagareishvili,R., Williams,T.,
Jackson,Y. and Bowers,Y.
TITLE     Endocrine Pancreas Consortium
JOURNAL   Unpublished (2000)
COMMENT   Contact: Douglas Melton, Klaus H. Kaestner, & Hiroshi Inoue
Endocrine Pancreas Consortium
Harvard University, Howard Hughes Medical Institute
Dept of Molecular and Cellular Biology, 7 Divinity Ave, Cambridge,
MA 02138
Tel: 617-495-1812
Fax: 617-495-8557
Email: dmelton@bioph.harvard.edu
Library was constructed by Dr. Douglas Melton DNA sequencing by:
Washington University Genome Sequencing Center This clone is
available royalty-free through LNU; please contact the IMAGE
consortium (info@image.lnl.gov) for further information
Seq primer: -40RP from Gibco
High quality sequence stop: 444.
FEATURES
SOURCE    location/Qualifiers
1..566
/organism="Mus musculus"
/db_xref="taxon:10090"
/clone_lib="IMAGE:5941344"
/clone_lib="Melton Amplified Mouse E16 5 Pancreas 3 M16S1
A"
/bex="Both"
/tissue_type="pancreas"
/dev_stage="Embryonic day E16.5"
/lab_host="DH10B"
/note="Vector: pSPORF1; Site_1: Not I; Site_2: Sal I;
Library constructed using Superscript plasmid library kit
(Life Technologies). cDNA made by oligo-dT priming.
Size-selected by column fractionation; average insert size
0.97 kb. Amplified once on solid support. cDNA library
Preparation: Guolin Chen."
BASE COUNT 97 a 190 c 153 g 126 t
ORIGIN
Alignment Scores:
Pred. No.: 0.426 Length: 566

```

Score: 95.50 Matches: 25  
Percent Similarity: 47.67% Conservative: 16  
Best Local Similarity: 29.07% Mismatches: 34  
Query Match: 17.30% Indels: 11  
DB: 14 Gaps: 1

US-10-020-441-2 (1-106) x BQ085897 (1-566)

Qy 4 LeuThrlLeuArgAenGlyAspGlnGlnGlyIleHisHisGlnProSerGlyValIysVal 23  
Db 270 CTTGCTGTGAAGAGAGGCCCCGAGCAGGTATCCCATCACTCGAGAGCTGTTTG 329  
Qy 24 AlaProGlyTyrMetProSerCysMetSerAlaArgGlnArgProIleAlaGlnThr 43  
Db 330 CATCACCTGTGACCGCGCTGTGTGGCTGT----- 362  
Qy 44 LeuSerGlyPheAspIleAlaIleValMetLeuAlaGlnMetValAsnLeuGlnSerGly 63  
Db 363 -----GGCAATGACATCGCCCTCATCAAGCTCTCAGCAGCGCCAGCTGGAGATGCC 416  
Qy 64 IleArgValIleSerLeuProGlnProSerAspIleProProGlyTyrGlyValPhe 83  
Db 417 GTCCAGCTGCGCTCATCTCCCGCTGTGATCTTCCACACAGCACCTGCTAC 476  
Qy 84 IleValGlyTyrGlyArg 89  
Db 477 ATCACCAGCTGGGCGCT 494

## RESULT 25

BMS11470 567 bp mRNA linear EST 15-FEB-2002  
LOCUS 1148f12.y1 Human insulinoma Homo sapiens cDNA clone IMAGE:5634214  
DEFINITION 5' similar to SW:EL3A\_HUMAN P09093 ELASTASE ITIA PRECURSOR ;, mRNA  
sequence.

ACCESSION BMS11470 GI:18682613  
VERSION BMS11470.1  
KEYWORDS EST.

SOURCE human.

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE 1 (bases 1 to 567)  
Melton,D., Brown,J., Kenty,G., Permutt,A., Lee,C., Kaestner,K.,  
Lemishka,I., Scarce,M., Brestelli,J., Gradwohl,G., Clifton,S.,  
Hillier,L., Marra,M., Pape,D., Wylie,T., Martin,J., Blistatn,A.,  
Schmitt,A., Theising,B., Ritter,E., Ronko,I., Bennett,J., Cardenas  
,M., Gibbons,M., McCann,R., Cole,R., Tsagarelshvili,R., Williams,T.,  
Jackson,Y. and Bowers,Y.

AUTHORS

TITLE Other ESTs: 1148f12.x1  
JOURNAL Contact: Douglas Melton, Klaus H. Kaestner, & Hiroshi Inoue  
COMMENT Endocrine Pancreas Consortium  
Harvard University, Howard Hughes Medical Institute  
Dept of Molecular and Cellular Biology, 7 Divinity Ave, Cambridge,  
MA 02138  
Tel: 617-495-1812  
Fax: 617-495-8557  
Email: dmelton@biom.harvard.edu  
Library was constructed by Dr. J. Ferrer in vivo mass-excised to  
pBluescript SK- by Dr. H. Inoue DNA sequencing by: Washington  
University Genome Sequencing Center For information on obtaining a  
clone please contact: Dr. Hiroshi Inoue (hinoue@im.wustl.edu)  
Seq primer: -40R from GIBCO  
High quality sequence stop: 430.

FEATURES  
source location/Qualifiers  
1..567  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone\_id="Human insulinoma"  
/clone\_lib="Human insulinoma"  
/tissue\_type="insulinoma"  
/lab\_host="DH10B (phage-resistant)"

/note="Organ: pancreas; Vector: pBluescript SK-; Site 1:  
XhoI; Site 2: EcoRI; Constructed with lambda ZAPII system  
(Stratagene) by Dr. J. Ferrer, in vivo mass-excised to  
pBluescript SK- by Dr. H. Inoue following the Washington  
University protocol  
(http://genome.wustl.edu/est/lambda\_protocol.shtml).  
Please contact Hiroshi Inoue, MD/PhD for further  
information on this library (Metabolism Division, Permutt  
Laboratory, Washington University School of Medicine, Box  
8127, 660 S Euclid Ave, St. Louis, MO 63110). Note: this  
is a Washington University Pancreas EST project library."

BASE COUNT 100 a 188 c 156 g 123 t

## ALIGNMENT SCORES:

Pred. No.: 0 426 Length: 567  
Score: 95.50 Matches: 25  
Percent Similarity: 47.67% Conservative: 16  
Best Local Similarity: 29.07% Mismatches: 34  
Query Match: 17.30% Indels: 11  
DB: 13 Gaps: 1

US-10-020-441-2 (1-106) x BMS11470 (1-567)

Qy 4 LeuThrlLeuArgAenGlyAspGlnGlnGlyIleHisHisGlnProSerGlyValIysVal 23  
Db 256 CTTGCTGTGAAGAGAGGCCCCGAGCAGGTATCCCATCACTCGAGAGCTGTTTG 315  
Qy 24 AlaProGlyTyrMetProSerCysMetSerAlaArgGlnArgProIleAlaGlnThr 43  
Db 316 CATCACCTGTGACCGCGCTGTGTGGCTGT----- 348  
Qy 44 LeuSerGlyPheAspIleAlaIleValMetLeuAlaGlnMetValAsnLeuGlnSerGly 63  
Db 349 -----GGCAATGACATCGCCCTCATCAAGCTCTCAGCAGCGCCAGCTGGAGATGCC 402  
Qy 64 IleArgValIleSerLeuProGlnProSerAspIleProProGlyTyrGlyValPhe 83  
Db 403 GTCCAGCTGCGCTCATCTCCCGCTGTGATCTTCCACACAGCACCTGCTAC 462  
Qy 84 IleValGlyTyrGlyArg 89  
Db 463 ATCACCAGCTGGGCGCT 480

RESULT 26  
BMS12185 571 bp mRNA linear EST 15-FEB-2002  
LOCUS 1176d06.y1 Human insulinoma Homo sapiens cDNA clone IMAGE:5636794  
DEFINITION 5' similar to SW:EL3A\_HUMAN P09093 ELASTASE ITIA PRECURSOR ;, mRNA  
sequence.

ACCESSION BMS12185 GI:18683328  
VERSION BMS12185.1  
KEYWORDS EST.

SOURCE human.

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE 1 (bases 1 to 571)  
Melton,D., Brown,J., Kenty,G., Permutt,A., Lee,C., Kaestner,K.,  
Lemishka,I., Scarce,M., Brestelli,J., Gradwohl,G., Clifton,S.,  
Hillier,L., Marra,M., Pape,D., Wylie,T., Martin,J., Blistatn,A.,  
Schmitt,A., Theising,B., Ritter,E., Ronko,I., Bennett,J., Cardenas  
,M., Gibbons,M., McCann,R., Cole,R., Tsagarelshvili,R., Williams,T.,  
Jackson,Y. and Bowers,Y.

AUTHORS

TITLE Other ESTs: 1176d06.x1  
JOURNAL Contact: Douglas Melton, Klaus H. Kaestner, & Hiroshi Inoue  
COMMENT Endocrine Pancreas Consortium  
Harvard University, Howard Hughes Medical Institute  
Dept of Molecular and Cellular Biology, 7 Divinity Ave, Cambridge,  
MA 02138  
Tel: 617-495-1812

FEATURES  
source location/Qualifiers  
1..567  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone\_id="Human insulinoma"  
/clone\_lib="Human insulinoma"  
/tissue\_type="insulinoma"  
/lab\_host="DH10B (phage-resistant)"

Fax: 617-495-8557  
 Email: dmelton@biohp.harvard.edu  
 Library was constructed by Dr. J. Ferrer in vivo mass-excised to  
 pBluescript SK- by Dr. H. Inoue DNA sequencing by: Washington  
 University Genome Sequencing Center for information on obtaining a  
 clone please contact: Dr. Hiroshi Inoue (hinoue@im.wustl.edu)  
 Seq primer: -40RP from Glibco  
 High quality sequence stop: 429.  
 Location/Qualifiers

# FEATURES

source

```

1..571
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:5636794"
/clone_lib="Human insulinoma"
/tissue_type="insulinoma"
/lab_host="DH10B (phage-resistant)"
/notes="organ: pancreas; Vector: pBluescript SK-; Site 1:
XhoI; Site 2: EcoRI; Constructed with lambda ZAPIT system
(Stratagene) by Dr. J. Ferrer, in vivo mass-excised to
pBluescript SK- by Dr. H. Inoue following the Washington
University protocol
(http://genome.wustl.edu/est/lambda_protocol.shtml).
Please contact Hiroshi Inoue, MD/PhD for further
information on this library (Metabolism Division, Permutt
Laboratory, Washington University School of Medicine, Box
8127, 660 S Euclid Ave, St. Louis, MO 63110). Note: this
is a Washington University Pancreas EST project library."
BASE COUNT    106 a    181 c    163 g    121 t
ORIGIN

```

## Alignment Scores:

Pred. No.:	0.43	Length:	571
Score:	95.50	Matches:	25
Percent Similarity:	47.67%	Conservative:	16
Best Local Similarity:	29.07%	Mismatches:	34
Query Match:	17.30%	Indels:	11
DB:	13	Gaps:	1

US-10-020-441-2 (1-106) x BM512185 (1-571)

```

QY 4 LeuThrLeuArgAsnGlyAspGlnGlnGlyLeuHisGlnProSerGlyValysVal 23
||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : : |||
Db 221 CTGTGCTGAAGAGGAGGCGCCGAGCAGGTGATCCCATCACTGAGAGAGCTGTTGG 280
QY 24 AlaProGlyTyrMetProSerCysMetSerAlaArgGlnArgProIleAlaGlnThr 43
||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : : |||
Db 281 CATCCACTTGGAACCGCTCGTGTGTGGCTGT----- 313
QY 44 LeuSerGlyPheAspIleAlaIleValMetLeuAlaGlnMetValAsnLeuGlnSerGly 63
||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : : |||
Db 314 -----GGCAATGACATCGCCCTCATCAAGCTCTCAGCAGCGCCAGCTGGAGATGCC 367
QY 64 IleArgValIleSerLeuProGlnProSerAspIleProProGlyThrGlyValPhe 83
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 368 GTCCAGCTGCCTCATCTCCCTCCCGCTGTGATCATCTTCCCAACAGACACCTGCTAC 427
QY 84 IleValGlyTyrGlyArg 89
||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : : |||
Db 428 ATCACCCTGCTGGGCGCT 445

```

## RESULT 27

BM967226 576 bp mRNA linear EST 29-APR-2002  
 LOCUS i332b02.y1 Melton Normalized Human Islet 4 N4-HIS 1 Homo sapiens  
 DEFINITION cDNA clone IMAGE:6136274 5' similar to SW:EI3A\_HUMAN F09093

ELASTAS IIIA PRECURSOR ;, mRNA sequence.  
 ACCESSION BM967226 GI:19561006  
 VERSION BM967226  
 KEYWORDS EST.

SOURCE human.  
 ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.

## REFERENCE

1 (bases 1 to 576)

AUTHORS

TITLE  
 JOURNAL  
 COMMENT

Other ESTs: i332b02.x1  
 Unpublished (2000)  
 Endocrine Pancreas Consortium  
 Contact: Douglas Melton, Klaus H. Kaestner, & Hiroshi Inoue  
 Endocrine Pancreas Consortium  
 Harvard University, Howard Hughes Medical Institute  
 Dept of Molecular and Cellular Biology, 7 Divinity Ave, Cambridge,  
 MA 02138  
 Tel: 617-495-1812  
 Fax: 617-495-8557  
 Email: dmelton@biohp.harvard.edu  
 Library was constructed by Dr. Douglas Melton DNA sequencing by:  
 Washington University Genome Sequencing Center for information on  
 obtaining a clone please contact: Juliana Brown  
 (brown@fas.harvard.edu)  
 Seq primer: -40RP from Glibco  
 High quality sequence stop: 446.  
 Location/Qualifiers

## FEATURES

source

```

1..576
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:6136274"
/clone_lib="Melton Normalized Human Islet 4 N4-HIS 1"
/sex="Both"
/tissue_type="Islets of Langerhans"
/dev_stage="Adult"
/lab_host="DH10B"
/notes="Organ: Pancreas; Vector: pSPORT1; Site 1: Not 1;
Site 2: Sat 1; Starting library constructed using
SuperScript Plasmid Library kit (Life Technologies). cDNA
made by oligo-dT priming. Size-selected by column
fractionation; average insert size 1.08 kb. Library was
amplified once on solid support and plasmid DNA from
library was prepared. The library DNA was normalized by
method #4 from Bonaldo, Lennon, and Soares 1996 Genome
Research 6:791-806; 0.5 microgram single-stranded library
plasmid DNA was mixed with 5 micrograms PCR product
representing library inserts and hybridized to an Ecot of
20. Single-stranded (unhybridized) plasmids were isolated
by hydroxyapatite chromatography and used to make this
library."
BASE COUNT    99 a    192 c    158 g    127 t
ORIGIN

```

## Alignment Scores:

Pred. No.:	0.435	Length:	576
Score:	95.50 <td>Matches:</td> <td>25</td>	Matches:	25
Percent Similarity:	47.67% <td>Conservative:</td> <td>16</td>	Conservative:	16
Best Local Similarity:	29.07% <td>Mismatches:</td> <td>34</td>	Mismatches:	34
Query Match:	17.30% <td>Indels:</td> <td>11</td>	Indels:	11
DB:	14	Gaps:	1

US-10-020-441-2 (1-106) x BM967226 (1-576)

```

QY 4 LeuThrLeuArgAsnGlyAspGlnGlnGlyLeuHisGlnProSerGlyValysVal 23
||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : : |||
Db 265 CTGTGCTGAAGAGGAGGCGCCGAGCAGGTGATCCCATCACTGAGAGAGCTGTTGG 324
QY 24 AlaProGlyTyrMetProSerCysMetSerAlaArgGlnArgProIleAlaGlnThr 43
||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : : |||
Db 325 CATCCACTTGGAACCGCTCGTGTGTGGCTGT----- 357
QY 44 LeuSerGlyPheAspIleAlaIleValMetLeuAlaGlnMetValAsnLeuGlnSerGly 63
||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : : |||
Db 358 -----GGCAATGACATCGCCCTCATCAAGCTCTCAGCAGCGCCAGCTGGAGATGCC 411

```

Qy 64 IleaYValIIeSerLeuProGlnProSerApIIeProProGlyThyGlyValphe 83  
 Db 412 GTCCAGCTCGCTCACTCCCGCTGGTGAATCCTTCCCAAGACACACCTGCTAC 471  
 Qy 84 IleaValGlyTYrGlyValy 89  
 Db 472 ATCACCAGCTGGGGCCGT 489

RESULT 28  
 BM967014 582 bp mRNA linear EST 29-APR-2002  
 LOCUS 1j79c10.y1 Melton Normalized Human Islet 4 N4-HIS 1 Homo sapiens  
 DEFINITION cDNA clone IMAGE:6135955 5' similar to SW:EL3A\_HUMAN P09093  
 ELASTASE I11A PRECURSOR ;, mRNA sequence.  
 ACCESSION BM967014  
 VERSION BM967014.1 GI:19560608  
 KEYWORDS EST.  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 1 (bases 1 to 582)  
 Melton, D., Brown, J., Kenty, G., Permutt, A., Lee, C., Kaestner, K.,  
 Lemishka, I., Scaerze, M., Brestelli, J., Gradwohl, G., Clifton, S.,  
 Hillier, L., Marra, M., Pape, D., Wylie, T., Martin, J., Blistrain, A.,  
 Schmitt, A., Theising, B., Rittler, E., Ronko, I., Bennett, J., Cardenas  
 , M., Gibbons, M., McCann, R., Cole, R., Tsagarishvili, R., Williams, T.,  
 Jackson, Y. and Bowers, Y.  
 Endocrine Pancreas Consortium  
 Unpublished (2000)  
 Other ESTs: 1j29c10.x1  
 Contact: Douglas Melton, Klaus H. Kaestner, & Hiroshi Inoue  
 Endocrine Pancreas Consortium  
 Harvard University, Howard Hughes Medical Institute  
 Dept of Molecular and Cellular Biology, 7 Divinity Ave, Cambridge,  
 MA 02138  
 Tel: 617-495-1812  
 Fax: 617-495-8557  
 Email: dmelton@biohp.harvard.edu  
 Library was constructed by Dr. Douglas Melton DNA sequencing by:  
 Washington University Genome Sequencing Center For information on  
 obtaining a clone please contact: Juliana Brown  
 (brown@fas.harvard.edu)  
 Seq primer: -40RP from Gibco  
 High quality sequence stop: 448.  
 Location/Qualifiers  
 1..582  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:6135955"  
 /clone\_1id="Melton Normalized Human Islet 4 N4-HIS 1"  
 /sex="Both"  
 /tissue\_type="Islets of Langerhans"  
 /dev\_stage="Adult"  
 /lab\_host="DH10B"  
 /note="Organ: Pancreas; Vector: pSPORn1; Site 1: Not 1;  
 Site 2: Sal 1; Starting library constructed using  
 SuperScript Plasmid library kit (Life Technologies). cDNA  
 made by oligo-dT priming. Size-selected by column  
 fractionation; average insert size 1.08 kb. Library was  
 amplified once on solid support and plasmid DNA from  
 library was prepared. The library DNA was normalized by  
 method #4 from Bonaldo, Lennon, and Sears 1996 Genome  
 Research 6:791-806; 0.5 microgram single-stranded library  
 plasmid DNA was mixed with 5 micrograms PCR product  
 representing library inserts and hybridized to an Ecot of  
 20. Single-stranded (unhybridized) plasmids were isolated  
 by hydroxyapatite chromatography and used to make this  
 library"

BASE COUNT 99 a 193 c 161 g 129 t  
 ORIGIN  
 Alignment Scores:

Pred. No.: 0.44 Length: 582  
 Score: 95.50 Matches: 25  
 Percent Similarity: 47.67% Conservative: 16  
 Best Local Similarity: 29.07% Mismatches: 34  
 Query Match: 17.30% Indels: 11  
 DB: 14 Gaps: 1

US-10-020-441-2 (1-106) x BM967014 (1-582)

Qy 4 LeuThrLeuAraGlnGlyAspGlnGlnGlyLeuHisGlnProSerGlyValIysVal 23  
 Db 265 CTTCGTGGAAGAGAGGCCCCGAGCATGATCCCATCAACTGTGAGAGCTGTTGTG 324  
 Qy 24 AlaProGlyTYrMetProSerCysMetSerAlaArgGlnArgGProIleAlaGlnThr 43  
 Db 325 CATCCACTCTGAACCCGCTGCTGTGGCTGT----- 357  
 Qy 44 LeuSerGlyPheAspIleAlaIleValMetLeuAlaGlnMetValAsnLeuGlnSerGly 63  
 Db 358 -----GGCAATGACATCGCCCTCATCAAGCTCTCACGACGAGCCAGCTGGAGATGCC 411  
 Qy 64 IleaYValIIeSerLeuProGlnProSerApIIeProProGlyThyGlyValphe 83  
 Db 412 GTCCAGCTCGCTCACTCCCGCTGGTGAATCCTTCCCAAGACACACCTGCTAC 471  
 Qy 84 IleaValGlyTYrGlyValy 89  
 Db 472 ATCACCAGCTGGGGCCGT 489

RESULT 29  
 BM512581 591 bp mRNA linear EST 15-FEB-2002  
 LOCUS 1j70f07.y1 Human insulinoma Homo sapiens cDNA clone IMAGE:5636484  
 DEFINITION 5' similar to SW:EL3A\_HUMAN P09093 ELASTASE I11A PRECURSOR ;, mRNA  
 sequence.  
 ACCESSION BM512581  
 VERSION BM512581.1 GI:18683724  
 KEYWORDS EST.  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 1 (bases 1 to 591)  
 Melton, D., Brown, J., Kenty, G., Permutt, A., Lee, C., Kaestner, K.,  
 Lemishka, I., Scaerze, M., Brestelli, J., Gradwohl, G., Clifton, S.,  
 Hillier, L., Marra, M., Pape, D., Wylie, T., Martin, J., Blistrain, A.,  
 Schmitt, A., Theising, B., Rittler, E., Ronko, I., Bennett, J., Cardenas  
 , M., Gibbons, M., McCann, R., Cole, R., Tsagarishvili, R., Williams, T.,  
 Jackson, Y. and Bowers, Y.  
 Endocrine Pancreas Consortium  
 Unpublished (2000)  
 Other ESTs: 1j70f07.x1  
 Contact: Douglas Melton, Klaus H. Kaestner, & Hiroshi Inoue  
 Endocrine Pancreas Consortium  
 Harvard University, Howard Hughes Medical Institute  
 Dept of Molecular and Cellular Biology, 7 Divinity Ave, Cambridge,  
 MA 02138  
 Tel: 617-495-1812  
 Fax: 617-495-8557  
 Email: dmelton@biohp.harvard.edu  
 Library was constructed by Dr. J. Ferrer in vivo mass-excised to  
 pBluescript SK- by Dr. H. Inoue DNA sequencing by: Washington  
 University Genome Sequencing Center For information on obtaining a  
 clone please contact: Dr. Hiroshi Inoue (hinoue@im.wustl.edu)  
 Seq primer: -40RP from Gibco  
 High quality sequence stop: 429.  
 Location/Qualifiers  
 1..591  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:5636484"  
 /clone\_1id="Human insulinoma"  
 /tissue\_type="Insulinoma"

FEATURES  
 source

```

/lab_host="DH10B (phage-resistant)"
/Note="Organ: pancreas; Vector: pluescript SK-; Site 1:
XhoI; Site 2: EcoRI; Constructed with lambda ZAPII system
(Stratagene) by Dr. J. Ferrer, in vivo mass-excised to
pluescript SK- by Dr. H. Inoue following the Washington
University protocol
(http://genome.wustl.edu/est/lambda_protocol.shtml).
Please contact Hiroshi Inoue, MD/PhD for further
information on this library (Metabolism Division, Permutt
Laboratory, Washington University School of Medicine, Box
8127, 660 S Euclid Ave, St. Louis, MO 63110). Note: this
is a Washington University Pancreas EST project library."
BASE COUNT      116 a      183 c      168 g      124 t
ORIGIN

Alignment Scores:
Pred. No.:      0.448      Length:      591
Score:          95.50      Matches:      25
Percent Similarity: 47.67%      Conservative: 16
Best Local Similarity: 29.07%      Mismatches: 34
Query Match:      17.30%      Indels:      11
DB:              13      Gaps:      1

US-10-020-441-2 (1-106) x BM512581 (1-591)

Qy      4      LeuThrLeuArgAsnGlyAspGlnGlnGlyLeuHisGlnProSerGlyValysVal 23
Db      42      CTTCGCTGTAAGAGAGGCGCCGACGAGTGATCCCATCACTGAGAGAGCTGTTTG 101
Qy      24      AlaProGlyTYrMetProSerCysMetSerAlaArgGlnArgProIleAlaGlnThr 43
Db      102      CATCACTCTGGAACCGCTGCTGTGTGCGCTGT----- 134
Qy      44      LeuSerGlyPheAspIleAlaIleValMetLeuAlaGlnMetValAsnLeuGlnSerGly 63
Db      135      -----GGCAATGACATCGCCCTCATCAAGCTTCCAGCAGCGCCAGCTGGAGATGCC 188
Qy      64      IleArgValIleSerLeuProGlnProSerAspIleProProProGlyTYrGlyAlaPhe 83
Db      189      GTCCAGCTGCCTCACTCCTCCCGCTGTGATCATCTTCCCAACAGACACCTGCTTAC 248
Qy      84      IleValGlyTYrGlyArg 89
Db      249      ATCACCAGCTGGGCGCGT 266

RESULT 30
LOCUS      BE969857      595 bp      mRNA      linear      EST 04-OCT-2000
DEFINITION      601679431P1 NIH_MGC_78 Homo sapiens cDNA clone IMAGE:394998 5',
ACCESSION      BE969857
VERSION      BE969857.1 GI:10582790
KEYWORDS      EST.
SOURCE      human.
ORGANISM      Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
1 (bases 1 to 595)
NIH-MGC http://mgc.nci.nih.gov/.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgarbs-remail.nih.gov
Tissue Procurement: CLONETECH Laboratories, Inc.
cDNA Library Preparation: CLONETECH Laboratories, Inc.
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Place: DLM815 row: m column: 07
High quality sequence stop: 593.
Location/Qualifiers
FEATURES

```

```

source
1. 595
/organism="Homo sapiens"
/db xref="taxon:9606"
/clone="IMAGE:394998"
/collection="NIH MGC 78"
/lab_host="DH10B (TI phage-resistant)"
/Note="Organ: pancreas; Vector: pNR-LIB (Clontech);
Site 1: SfiI (ggccatcgcc); Site 2: SfiI (ggccatcgcc
); 5' and 3' adaptors were used in cloning as follows: 5'
adaptor sequence: 5'-CACGGCATTAATGGCC-3' and 3' adaptor
sequence: 5'-ATTCTAGAGCGCGAGCGGCGGACATG-dt(30)BN-3'
(where B = A, C, or G and N = A, C, G, or T). Average
insert size 1.2 kb (range 0.5-4.0 kb). 14/15 colonies
contained inserts by PCR. This library was enriched for
full-length clones and was constructed by Clontech
Laboratories (Palo Alto, CA)."
BASE COUNT      104 a      199 c      165 g      127 t
ORIGIN

Alignment Scores:
Pred. No.:      0.452      Length:      595
Score:          95.50      Matches:      25
Percent Similarity: 47.67%      Conservative: 16
Best Local Similarity: 29.07%      Mismatches: 34
Query Match:      17.30%      Indels:      11
DB:              12      Gaps:      1

US-10-020-441-2 (1-106) x BE969857 (1-595)

Qy      4      LeuThrLeuArgAsnGlyAspGlnGlnGlyLeuHisGlnProSerGlyValysVal 23
Db      263      CTTCGCTGTAAGAGAGGCGCCGACGAGTGATCCCATCACTGAGAGAGCTGTTTG 322
Qy      24      AlaProGlyTYrMetProSerCysMetSerAlaArgGlnArgProIleAlaGlnThr 43
Db      323      CATCACTCTGGAACCGCTGCTGTGTGCGCTGT----- 355
Qy      44      LeuSerGlyPheAspIleAlaIleValMetLeuAlaGlnMetValAsnLeuGlnSerGly 63
Db      356      -----GGCAATGACATCGCCCTCATCAAGCTTCCAGCAGCGCCAGCTGGAGATGCC 409
Qy      64      IleArgValIleSerLeuProGlnProSerAspIleProProProGlyTYrGlyAlaPhe 83
Db      410      GTCCAGCTGCCTCACTCCTCCCGCTGTGATCATCTTCCCAACAGACACCTGCTTAC 469
Qy      84      IleValGlyTYrGlyArg 89
Db      470      ATCACCAGCTGGGCGCGT 487

RESULT 31
LOCUS      BM310730      614 bp      mRNA      linear      EST 03-JAN-2002
DEFINITION      i947905.y1 HR85 test Homo sapiens cDNA 5' similar to SW-EL3A_HUMAN
ACCESSION      BM310730
VERSION      BM310730.1 GI:18045075
KEYWORDS      EST.
SOURCE      human.
ORGANISM      Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
1 (bases 1 to 614)
Melton, D., Brown, J., Kenty, G., Permutt, A., Lee, C., Kaestner, K.,
Lemishka, I., Scearce, M., Bresnelli, J., Gradow, G., Clifton, S.,
Hillier, L., Marra, M., Page, D., Wylie, T., Martin, J., Bissain, A.,
Schmitt, A., Thersing, B., Ritter, E., Konko, I., Bennett, J., Gardenas
, M., Gibbons, M., McCann, R., Cole, R., Tsagarelis, R., Williams, T.,
Jackson, Y., and Bowers, R.
Endocrine Pancreas Consortium
Unpublished (2000)
Other ESTs: i947905.x1
Contact: Douglas Melton, Klaus H. Kaestner, & Hiroshi Inoue
Endocrine Pancreas Consortium

```





JOURNAL Unpublished (2000)  
 COMMENT Contact: Douglas Melton, Klaus H. Kaestner, & Hiroshi Inoue  
 Endocrine Pancreas Consortium  
 Harvard University, Howard Hughes Medical Institute  
 Dept of Molecular and Cellular Biology, 7 Divinity Ave, Cambridge,  
 MA 02138  
 Tel: 617-495-1812  
 Fax: 617-495-8557  
 Email: dmelton@iobp.harvard.edu  
 Library was constructed by Dr. Hiroshi Inoue DNA sequencing by:  
 Washington University Genome Sequencing Center For information on  
 obtaining a clone please contact: Dr. Hiroshi Inoue  
 (hinoue@im.wustl.edu)  
 Seq primer: -40UP from Gdbco  
 High quality sequence stop: 359.  
 Location/Qualifiers  
 1. 629  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone\_lib="HR85\_islet"  
 /issue\_type="Purified pancreatic islet"  
 /lab\_host="DH10B"  
 /note="Organ: Pancreas; Vector: pBluescript SK(-); Site\_1:  
 NotI; Site\_2: XhoI; cDNA made by oligo-dT priming.  
 Size-selected on agarose gel. Average insert size ~1kb. 5'  
 XhoI site was destroyed after directional cloning.  
 Amplified once. Contact information: Hiroshi Inoue, MD,  
 Metabolism Div. (Alan Permutt Lab), Washington University  
 School of Medicine, Box 8127, 660 South Euclid Ave., St.  
 Louis, MO 63110, E-mail: hinoue@ingate.wustl.edu, Tel:  
 314-362-1916, Fax: 314-747-2692."  
 BASE COUNT 123 a 172 c 197 g 136 t 1 others  
 ORIGIN  
 Alignment Scores:  
 Pred. No.: 0.484 Length: 629  
 Score: 95.50 Matches: 25  
 Percent Similarity: 47.62% Conservative: 15  
 Best Local Similarity: 29.76% Mismatches: 33  
 Query Match: 17.30% Indels: 11  
 DB: 13 Gaps: 1  
 US-10-020-441-2 (1-106) x BM272242 (1-629)  
 Oy 6 LeuAgaAGlAAspGlnGlyIleHISGlnProSerGlyValLysValA1aPro 25  
 Db 617 GTAAGGAGAGGAGCCCGAGAGGTATCCCATCACTCGGGAGCCTTTGTGATCCA 558  
 Oy 26 GlyTYrMetProSerCyMetSerAlaArgGlnArgArgProIleAglInThrLeuSer 45  
 Db 557 CTCTGAGACCCGCTCGTGTGTGGCTGT----- 531  
 Oy 46 GlyPheAspIleAlaIleValMetLeuAlaGlnMetValAsnLeuGlnSerGlyIleArg 65  
 Db 530 GGCAATGACATGACCTCTCAAGCTTCACGACGCGCCACCTGGAGACCGCGCCAG 471  
 Oy 66 ValIleSerLeuProGlnProSerAspIleProProGlyTYrGlyValPheIleVal 85  
 Db 470 CTCGCTTACCTCTCCCGTGTGATCTTCCACAGAGACCCCTGCTATCACC 411  
 Oy 86 GlyTYrGlyArg 89  
 Db 410 GGCTGGGGCGGT 399  
 RESULT 34  
 BQ777957  
 LOCUS BQ777957 632 bp mRNA linear EST 26-JUN-2002  
 DEFINITION 1139b11.y1 HR85 islet Homo sapiens cDNA clone IMAGE:6032228 5'  
 similar to SW:ELJA\_HUMAN P09093 ELASTASE IITA PRECURSOR ;, mRNA  
 sequence.  
 ACCESSION BQ777957  
 VERSION BQ777957.1 GI:21986429  
 KEYWORDS EST.

SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 REFERENCE 1 (bases 1 to 632)  
 AUTHORS Melton,D., Brown,J., Kenty,G., Permutt,A., Lee,C., Kaestner,K.,  
 Lemishka,I., Scaerza,M., Brestelli,J., Gradwohl,G., Clifton,S.,  
 Hillier,L., Marra,M., Pape,D., Wylie,T., Martin,J., Blisstein,A.,  
 Schmitt,A., Theising,B., Ritzer,E., Ronko,I., Bennett,J., Cardenas  
 , M., Gibbons,M., McCann,R., Cole,R., Tsagarishvili,R., Williams,T.,  
 Jackson,X. and Bowers,Y.  
 Endocrine Pancreas Consortium  
 Unpublished (2000)  
 Other ESTs: 1139b11.x1  
 Contact: Douglas Melton, Klaus H. Kaestner, & Hiroshi Inoue  
 Endocrine Pancreas Consortium  
 Harvard University, Howard Hughes Medical Institute  
 Dept of Molecular and Cellular Biology, 7 Divinity Ave, Cambridge,  
 MA 02138  
 Tel: 617-495-1812  
 Fax: 617-495-8557  
 Email: dmelton@iobp.harvard.edu  
 Library was constructed by Dr. Hiroshi Inoue DNA sequencing by:  
 Washington University Genome Sequencing Center For information on  
 obtaining a clone please contact: Dr. Hiroshi Inoue  
 (hinoue@im.wustl.edu)  
 Seq primer: -40RP from Gdbco  
 High quality sequence stop: 460.  
 Location/Qualifiers  
 1. 632  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone\_lib="IMAG6:6032228"  
 /issue\_type="Purified pancreatic islet"  
 /lab\_host="DH10B"  
 /note="Organ: Pancreas; Vector: pBluescript SK(-); Site\_1:  
 NotI; Site\_2: XhoI; cDNA made by oligo-dT priming.  
 Size-selected on agarose gel. Average insert size ~1kb. 5'  
 XhoI site was destroyed after directional cloning.  
 Amplified once. Contact information: Hiroshi Inoue, MD,  
 Metabolism Div. (Alan Permutt Lab), Washington University  
 School of Medicine, Box 8127, 660 South Euclid Ave., St.  
 Louis, MO 63110, E-mail: hinoue@ingate.wustl.edu, Tel:  
 314-362-1916, Fax: 314-747-2692."  
 BASE COUNT 114 a 204 c 177 g 137 t  
 ORIGIN  
 Alignment Scores:  
 Pred. No.: 0.486 Length: 632  
 Score: 95.50 Matches: 25  
 Percent Similarity: 47.67% Conservative: 16  
 Best Local Similarity: 29.07% Mismatches: 34  
 Query Match: 17.30% Indels: 11  
 DB: 14 Gaps: 1  
 US-10-020-441-2 (1-106) x BQ777957 (1-632)  
 Oy 4 LeuThrLeuAgaAGlAAspGlnGlyIleHISGlnProSerGlyValLysVal 23  
 Db 240 CTGTGTGAGAGAGAGGAGCCCGAGAGGTATCCCATCACTGAGAGAGCTTTGG 299  
 Oy 24 AlaProGlyTYrMetProSerCyMetSerAlaArgGlnArgArgProIleAglInThr 43  
 Db 300 CATCACTCTGGAACCGCTCGTGTGTGGCTGT----- 332  
 Oy 44 LeuSerGlyPheAspIleAlaIleValMetLeuAlaGlnMetValAsnLeuGlnSerGly 63  
 Db 333 -----GGCAATGACATGACCTCTCAAGCTTCACGACGCGCCACCTGGAGATGCC 386  
 Oy 64 IleArgValIleSerLeuProGlnProSerAspIleProProGlyTYrGlyValPhe 83  
 Db 387 GTCCAGCTCGCTCACTCTCCCGTGTGATCTTCCACAGAGACCCCTGCTAC 446

QY 84 11evAlG1YrG1YArG 89  
 DB 447 ATCACCgGCTGGGCGCT 464

RESULT 35  
 BI711146/c 639 bp mRNA linear EST 11-MAR-2002  
 LOCUS id92c05.x1 Human insulinoma Homo sapiens cDNA clone IMAGE:5023112  
 DEFINITION 3' similar to SW:ELJA\_HUMAN P09093 ELASTASE IITA PRECURSOR ;, mRNA sequence.

ACCESSION BI711146 GI:15686834  
 VERSION BI711146  
 KEYWORDS EST.  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 1 (bases 1 to 639)  
 Melton,D., Brown,J., Keny,G., Permutt,A., Lee,C., Kaestner,K., Hillier,L., Marra,M., Page,D., Wylie,T., Martin,J., Bistrain,A., Schmitt,A., Theising,B., Ritter,E., Ronko,I., Bennett,J., Cardenas,M., Gibbons,M., McCann,R., Cole,R., Tsagarashvili,R., Williams,T., Jackson,Y. and Bowers,Y.  
 'M., Gibbons,M., McCann,R., Cole,R., Tsagarashvili,R., Williams,T., Jackson,Y. and Bowers,Y.  
 Endocrine Pancreas Consortium  
 Unpublished (2000)  
 Other ESTs: id92c05.y1  
 Contact: Douglas Melton, Klaus H. Kaestner, & Hiroshi Inoue  
 Endocrine Pancreas Consortium  
 Harvard University, Howard Hughes Medical Institute  
 Dept of Molecular and Cellular Biology, 7 Divinity Ave, Cambridge, MA 02138  
 Tel: 617-495-1812  
 Fax: 617-495-8557  
 Email: dmelton@bioh.harvard.edu  
 Library was constructed by Dr. J. Ferrer in vivo mass-excised to pBluescript SK- by Dr. H. Inoue DNA sequencing by: Washington University Genome Sequencing Center For information on obtaining a clone please contact: Dr. Hiroshi Inoue (hinoue@im.wustl.edu)  
 Seq primer: -40UP from Gldco  
 High quality sequence stop: 481.  
 Location/Qualifiers  
 1..639  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:5023112"  
 /clone\_id="Human insulinoma"  
 /tissue\_type="Insulinoma"  
 /lab\_host="DH10B (phage-resistant)"  
 /note="Organ: pancreas; Vector: pBluescript SK-. Site 1: XhoI; Site 2: EcoRI; Constructed with lambda ZapII system (Stratagene) by Dr. J. Ferrer, in vivo mass-excised to pBluescript SK- by Dr. H. Inoue following the Washington University protocol (http://genome.wustl.edu/est/lambda\_protocol.shtml).  
 Please contact Hiroshi Inoue, MD/PhD for further information on this library (Metabolism Division, Permutt Laboratory, Washington University School of Medicine, Box 8127, 660 S Euclid Ave, St. Louis, MO 63110). Note: this is a Washington University Pancreas EST project library."

BASE COUNT 129 a 171 c 192 g 147 t  
 ORIGIN

Alignment Scores:  
 Pred. No.: 0.493 Length: 639  
 Score: 95.50 Matches: 25  
 Percent Similarity: 47.67% Conservative: 16  
 Best Local Similarity: 29.07% Mismatches: 34  
 Query Match: 17.30% Indels: 11  
 DB: 13 Gaps: 1

US-10-020-441-2 (1-106) x BI711146 (1-639)

QY 4 LeuThrLeuArGAsnG1YAspGlnGlnG1YleH1sHsGlnProSerG1YValYsVal 23  
 DB 627 CTTGCTGTGAAGAGAGGCCCCGAGAGGTGATCCCCATCAACTCGAGAGAGCTGTTGTG 568  
 QY 24 AlaProG1YrMetProSerCyMetSerAlaargGlnArGArGProIleAlaGlnThr 43  
 DB 567 CATCAGCTCTGGAACCGCTCGGTGTGTGCTGT----- 535  
 QY 44 LeuSerG1YpHeAspIleAlaIleValMetLeuAlaGlnMetValAsnLeuGlnSerG1Y 63  
 DB 534 -----GGCAATGACATGACCCCTCATGACTCTCAGCGCCAGCCGAGATGCC 481  
 QY 64 11evAl111eserLeuProGlnProSerAspIleProProG1YrHrG1YAlphe 83  
 DB 480 GTCCAGCTCGCTCATCTCCCTCCGCTGTGATCATCTTCCCAACACACCTGCTAC 421

QY 84 11evAlG1YrG1YArG 89  
 DB 420 ATCACCgGCTGGGCGCT 403

RESULT 36  
 AM583424/c 758 bp mRNA linear EST 13-MAR-2002  
 LOCUS ia02e09.x1 Human Pancreatic Islets Homo sapiens cDNA clone IMAGE:5637160 3' similar to gb:M18692 ELASTASE IIB PRECURSOR (HUMAN); gb:X04573 Mouse mRNA for preproelastase II (MOUSE);, mRNA sequence.

ACCESSION AM583424  
 VERSION AM583424.1 GI:7260382  
 KEYWORDS EST.  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 1 (bases 1 to 758)  
 Melton,D., Meadows,A., Clifton,S., Hillier,L., Marra,M., Page,D., Wylie,T., Martin,J., Bistrain,A., Schmitt,A., Theising,B., Ritter,E., Ronko,I., Bennett,J., Cardenas,M., Gibbons,M., McCann,R., Cole,R., Tsagarashvili,R., Williams,T., Jackson,Y. and Bowers,Y.  
 Maahu-Harvard Pancreas EST Project  
 Unpublished (2000)  
 Contact: Douglas Melton, Klaus H. Kaestner, & Hiroshi Inoue  
 Endocrine Pancreas Consortium  
 Harvard University, Howard Hughes Medical Institute  
 Dept of Molecular and Cellular Biology, 7 Divinity Ave, Cambridge, MA 02138  
 Tel: 617-495-1812  
 Fax: 617-495-8557  
 Email: dmelton@bioh.harvard.edu  
 Libraries were constructed by Dr. Douglas Melton  
 DNA sequencing by: Washington University Genome Sequencing Center For information on obtaining a clone please contact: Juliana Brown (brown@fas.harvard.edu)  
 This sequence now available from the IMAGE consortium, for clone orders contact: info@image.llnl.gov  
 Seq primer: -40UP from Gldco  
 High quality sequence stop: 412.  
 Location/Qualifiers  
 1..758  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:5637160"  
 /clone\_id="Human Pancreatic Islets"  
 /tissue\_type="Islets of Langerhans"  
 /dev\_stage="Adult"  
 /lab\_host="DH10B"  
 /note="Organ: Pancreas; Vector: pSPORT1; Site 1: Not 1; Site 2: Sal 1, library constructed using Superscript Plasmid Library Kit (Life Technologies). cDNA made by oligo-dT priming. Size-selected by column fractionation; average insert size 1.08 kb. Primary library, unamplified."

FEATURES  
 source

BASE COUNT 151 a 213 c 232 g 161 t 1 others  
ORIGIN

## Alignment Scores:

Pred. No.: 0.606 Length: 758  
Score: 95.50 Matches: 27  
Percent Similarity: 48.24% Conservative: 14  
Best Local Similarity: 31.76% Mismatches: 32  
Query Match: 17.30% Indels: 12  
DB: 10 Gaps: 1

US-10-020-441-2 (1-106) x AWS83424 (1-758)

Qy 6 LeuArgaenGIYAspGInGInGly-1LeHshISgInProSerGIYValIyValAlaPr 25  
Db 612 GTAAAGAGAGGGCCCCCAAGGGTGATCCCCCAACTGGGGCCTCTTTGTGCATCC 553  
Qy 25 cGIYTYrMeTProSerCyMeTSeRaIaArgInaArgProIIeAlaInThrLeuSe 45  
Db 552 ACTCGGAGCCCGCTGTGTGTGGCTGT----- 525  
Qy 45 rGIYpheaSPIleAlaIleValMeTleuAlaGInMeTValaSnLeuGInSeRGIYIleAr 65  
Db 524 -GGCATATGACATGCGCTCATCAAGCTCTACAGCAGCCCGCCTGGAGACCGNGTCA 466  
Qy 65 vAlIleSeRleuProGInProSeRaSPIleProProProGIYThrGIYValpHeIleVa 85  
Db 465 GCTCGCCCTACCTCCCTCGGCTGGTGAATCTCTCCCAACAGACACCTGTATCATCAC 406  
Qy 85 lGIYTYrGIYArg 89  
Db 405 CGGCTGGGGCCGT 393

RESULT 37  
BE969722 828 bp mRNA linear EST 04-OCT-2000  
LOCUS 601679672P1 NIH\_MGC\_78 Homo sapiens cDNA clone IMAGE:3950081 5',  
DEFINITION mRNA sequence.  
ACCESSION BE969722  
VERSION BE969722.1 GI:10582655  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE NIH-MGC http://mgc.nci.nih.gov/.  
1 (bases 1 to 828)  
AUTHORS National Institutes of Health, Mammalian Gene Collection (MGC)  
TITLE Unpublished (1999)  
JOURNAL  
COMMENT Contact: Robert Strausberg, Ph.D.  
Email: cgabbs-remail.nih.gov  
Tissue Procurement: CLONETECH Laboratories, Inc.  
cDNA Library Preparation: CLONETECH Laboratories, Inc.  
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
DNA Sequencing by: Incyte Genomics, Inc.  
Clone distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL at:  
http://image.llnl.gov  
Plate: L1CM815 row: p column: 18  
High quality sequence stop: 634.

## FEATURES

source 1..828  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:3950081"  
/clone\_1ib="NIH\_MGC\_78"  
/lab\_host="DH10B (TI phage-resistant)"  
/note="Organ: pancreas; Vector: pDNR-LIB (Clontech);  
Site 1: SfiI (ggcgagctcgcc); Site 2: SfiI (ggcgatagggc  
); 5' and 3' adaptors were used in cloning as follows: 5'  
adaptor sequence: 5'-CACGGCCATTATGGCC-3' and 3' adaptor  
sequence: 5'-ATCTAGAGCGCGGCGGACATG-dt(30)BN-3'  
(where B = A, C, or G and N = A, C, G, or T). Average

insert size 1.2 kb (range 0.5-4.0 kb). 14/15 colonies  
contained inserts by PCR. This library was enriched for  
full-length clones and was constructed by Clontech  
Laboratories (Palo Alto, CA)."

BASE COUNT 167 a 243 c 253 g 165 t  
ORIGIN

## Alignment Scores:

Pred. No.: 0.675 Length: 828  
Score: 95.50 Matches: 25  
Percent Similarity: 47.67% Conservative: 16  
Best Local Similarity: 29.07% Mismatches: 34  
Query Match: 17.30% Indels: 11  
DB: 12 Gaps: 1

US-10-020-441-2 (1-106) x BE969722 (1-828)

Qy 4 LeuThrLeuAgaSnGIYAspGInGInGly-1LeHshISgInProSerGIYValIyVal 23  
Db 264 CTGCTGTGAAGAGGGCCCCGAGCAGGTGATCCCATCACTGAGAGCTGTTTGTG 323  
Qy 24 AlAProGIYTYrMeTProSerCyMeTSeRaIaArgInaArgProIIeAlaInThr 43  
Db 324 CATCCACTCTGGAACCGCTCGTGTGTGGCTGT----- 356  
Qy 44 LeuSeRGIYpheaSPIleAlaIleValMeTleuAlaGInMeTValaSnLeuGInSeRGIY 63  
Db 357 -----GGCATATGACATGCGCTCATCAAGCTCTACAGCAGCCCGCCTGGAGATGCC 410  
Qy 64 lIeArGIYAlIleSeRleuProGInProSeRaSPIleProProProGIYThrGIYValpHe 83  
Db 411 GTCCAGCTCGCTCATCTCCCTCGGCTGGTGAATCTCTCCCAACAGACACCTGTATC 470  
Qy 84 lIeValGIYTYrGIYArg 89  
Db 471 ATCACCAGCTGGGGCCGT 488

RESULT 38  
BE969613 854 bp mRNA linear EST 04-OCT-2000  
LOCUS 601679743P1 NIH\_MGC\_78 Homo sapiens cDNA clone IMAGE:3949834 5',  
DEFINITION mRNA sequence.  
ACCESSION BE969613  
VERSION BE969613.1 GI:10582546  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE NIH-MGC http://mgc.nci.nih.gov/.  
1 (bases 1 to 854)  
AUTHORS National Institutes of Health, Mammalian Gene Collection (MGC)  
TITLE Unpublished (1999)  
JOURNAL  
COMMENT Contact: Robert Strausberg, Ph.D.  
Email: cgabbs-remail.nih.gov  
Tissue Procurement: CLONETECH Laboratories, Inc.  
cDNA Library Preparation: CLONETECH Laboratories, Inc.  
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
DNA Sequencing by: Incyte Genomics, Inc.  
Clone distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL at:  
http://image.llnl.gov  
Plate: L1CM815 row: f column: 11  
High quality sequence stop: 660.

## FEATURES

source 1..854  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:3949834"  
/clone\_1ib="NIH\_MGC\_78"  
/lab\_host="DH10B (TI phage-resistant)"  
/note="Organ: pancreas; Vector: pDNR-LIB (Clontech);  
Site 1: SfiI (ggcgagctcgcc); Site 2: SfiI (ggcgatagggc



```

/clone_11b="NIH_MGC_122"
/lab_host="DH10B"
/note="Organ: pooled lung and spleen; Vector: PCMV-SPORT6;
Site_1: NotI; Site_2: EcoRV (destroyed); RNA source
anonymous pool of 24 week female lung, 16 week female
spleen, and 20-22 week male spleens. Library is oligo-dT
primed and directionally cloned (EcoRV site is destroyed
upon cloning). Average insert size 1.4 kb, insert size
range 1-3 kb. Library is normalized and enriched for
full-length clones and was constructed by C. Gruber
(Invitrogen). Research Genetics tracking code 026. Note:
this is a NIH_MGC Library."
BASE COUNT      239 a      352 c      344 g      227 t
ORIGIN

```

```

Alignment Scores:
Pred. No.:      1.02      Length:      1162
Score:          95.50      Matches:      25
Percent Similarity: 47.67%      Conservative: 16
Best Local Similarity: 29.07%      Mismatches: 34
Query Match:      17.30%      Indels:      11
DB:              14      Gaps:      1

```

US-10-020-441-2 (1-106) x BM926658 (1-1162)

```

QY      4  leuThrLeuArgAsnGlyAspGlnGlnGlyIleHisGlnProSerGlyValIysVal 23
      |||  :::::  |||  :::::  |||  :::::  |||  :::::  |||  :::::  |||
Db      253 CTTGCTGTGAAGAGGCGCCCGACGACGATGCCCATCACTCGAGAGCTGTTGTG 312

QY      24  AlaProGlyTyrMetProSerCysMetSerAlaArgGlnArgArgProIleAlaGlnThr 43
      |||  :::::  |||  :::::  |||  :::::  |||  :::::  |||  :::::  |||
Db      313 CATCACTCTGGAACCGCTCGTGTGGCTGT----- 345

QY      44  leuSerGlyPheAspIleAlaIleValMetLeuAlaGlnMetValAsnLeuGlnSerGly 63
      |||  :::::  |||  :::::  |||  :::::  |||  :::::  |||  :::::  |||
Db      346 -----GGCAATGACATCGCCCTCATCAAGCTCTCAAGCGAGCCCGAGCTGGAGATGCC 399

QY      64  IleArgValIleSerLeuProGlnProSerAspIleProProGlyTyrGlyValPhe 83
      :::::  |||  :::::  |||  :::::  |||  :::::  |||  :::::  |||  :::::  |||
Db      400 GTCCAGCTCGCTCACTCCCTCCCGCTGTGTGATCTCTCCCAACAAGACACCTGCTAC 459

QY      84  IleValGlyTyrGlyArg 89
      |||  :::::  |||  :::::  |||  :::::  |||  :::::  |||  :::::  |||
Db      460 ATCACCGGCTGGGCGCGT 477

```

Search completed: April 8, 2003, 19:52:07  
 Job time : 1482.44 secs

GenCore version 5.1.4 p5.4578  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 9, 2003, 12:39:51 ; Search time 13.377 Seconds  
(without alignments)  
159.378 Million cell updates/sec

Title: US-10-020-441-3

Perfect score: 90

Sequence: 1 VGYGRDNDNDPDSRKN 16

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :  
1: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.\*  
2: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.\*  
3: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.\*  
4: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.\*  
5: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.\*  
6: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.\*  
7: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.\*  
8: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.\*  
9: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.\*  
10: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.\*  
11: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.\*  
12: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.\*  
13: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.\*  
14: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.\*  
15: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.\*  
16: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.\*  
17: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.\*  
18: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.\*  
19: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.\*  
20: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.\*  
21: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.\*  
22: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.\*  
23: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2002.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length DB	ID	Description
1	90	100.0	16	21	AA67588
2	90	100.0	106	21	AA67587
3	55	61.1	11	16	AAW21484
4	47	52.2	757	22	AAU19569
5	47	52.2	757	23	ABP51294
6	47	52.2	762	23	ABG60124
7	46	51.1	1429	23	ABP47797
8	45	50.0	149	21	AA10642
9	45	50.0	717	21	AA646818
10	45	50.0	1224	22	ABG15848

11	44	48.9	614	23	AAU09996
12	44	48.9	614	23	AA678515
13	43	47.8	84	22	AAU43208
14	43	47.8	229	22	AAW40752
15	43	47.8	241	22	AA682400
16	43	47.8	258	23	ABP39919
17	43	47.8	297	22	ABG10604
18	43	47.8	444	22	ABB68139
19	43	47.8	451	21	AA94944
20	43	47.8	477	22	AAW38966
21	43	47.8	515	22	ABG10605
22	43	47.8	809	22	ABG10607
23	42	46.7	205	22	ABG29958
24	42	46.7	215	22	ABB61193
25	42	46.7	254	22	ABG29954
26	42	46.7	361	22	ABG29956
27	42	46.7	500	22	AAU16409
28	42	46.7	1005	21	AAW43890
29	42	46.7	1179	23	ABW31867
30	42	46.7	1382	18	AAW31867
31	42	46.7	2559	23	ABW57181
32	41	45.6	127	20	AAV59894
33	41	45.6	127	22	AAU87108
34	41	45.6	168	21	AAW29469
35	41	45.6	223	22	AAU87169
36	41	45.6	229	22	AAU17282
37	41	45.6	310	23	AAW15538
38	41	45.6	376	21	AAW36931
39	41	45.6	415	21	AAW36930
40	41	45.6	448	21	AAW36929
41	41	45.6	514	22	AAW48240
42	41	45.6	572	22	ABW57850
43	41	45.6	789	23	ABW7331
44	41	45.6	854	22	AAW67391
45	41	45.6	952	22	ABW69760

## ALIGNMENTS

RESULT 1	
AA67588	
ID	AA67588 standard; Protein; 16 AA.
XX	
AC	AA67588;
XX	
DT	19-JUN-2000 (first entry)
XX	
DE	S. mansoni cercarial elastase protein fragment.
XX	
KW	Schistosoma parasite; fusion protein; cercarial elastase; vaccine;
KW	antibody response; schistosomicide.
XX	
OS	Schistosoma mansoni.
XX	
PN	EP992582-A2.
XX	
PD	12-APR-2000.
XX	
PF	05-OCT-1999; 99EP-0307832.
XX	
PR	07-OCT-1998; 98GB-0021821.
XX	
PA	(UYWA-) UNIV WALEIS BANGOR.
XX	
PI	Doenhoff M, Sayers J;
XX	
DR	WPI: 2000-259136/23.
DR	N-PSDB; AA290608.
XX	
PT	New vaccine for treatment of Schistosoma infections contains a recombinant fusion protein comprising cercarial elastase sequence fused to bacterial, phage or viral protein -

XX Claim 5; Page 18; 26pp; English.  
 PS  
 CC The invention provides a vaccine comprising a recombinant fusion protein  
 CC capable of eliciting immunity against Schistosoma parasites. The fusion  
 CC protein comprises the 27 or 28 kDa cercarial elastase sequence of  
 CC S. mansoni or an active fragment, homologue or variant, fused to a  
 CC bacterial, phage or viral protein. The vaccine containing the fusion  
 CC protein has been found to induce a significantly increased antibody  
 CC response against schistosoma infections, compared to the use of S.  
 CC mansoni cercarial elastase in its native form. The present sequence  
 CC represents the S. mansoni cercarial elastase protein fragment encoded by  
 CC exon 1 of the cercarial elastase gene, used in the fusion protein of the  
 CC invention.  
 XX  
 SQ Sequence 16 AA;  
 Query Match 100.0%; Score 90; DB 21; Length 16;  
 Best Local Similarity 100.0%; Pred. No. 2e-08;  
 Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 VGYGRDNDNDPSPSRKN 16  
 DB 1 VGYGRDNDNDPSPSRKN 16  
 RESULT 2  
 ID AAY67587 standard; Protein; 106 AA.  
 AC AAY67587;  
 XX  
 DT 19-JUN-2000 (first entry)  
 XX  
 DE S. mansoni cercarial elastase protein fragment.  
 KW Schistosoma parasite; fusion protein; cercarial elastase; vaccine;  
 KM antibody response; schistosomicide.  
 XX  
 OS Schistosoma mansoni.  
 XX  
 PN EP992582-A2.  
 XX  
 PD 12-APR-2000.  
 XX  
 PF 05-OCT-1999; 99EP-0307832.  
 XX  
 PR 07-OCT-1998; 98GB-0021821.  
 XX  
 PA (UYWA-) UNIV WALES BANGOR.  
 XX  
 PI Doenhoff M, Sayers J;  
 XX  
 DR WPI; 2000-259136/23.  
 DR N-PSDB; AAZ90608.  
 XX  
 PT New vaccine for treatment of Schistosoma infections contains a  
 PT recombinant fusion protein comprising cercarial elastase sequence fused  
 PT to bacterial, phage or viral protein -  
 XX  
 Claim 3; Page 17; 26pp; English.  
 PS  
 CC The invention provides a vaccine comprising a recombinant fusion protein  
 CC capable of eliciting immunity against Schistosoma parasites. The fusion  
 CC protein comprises the 27 or 28 kDa cercarial elastase sequence of  
 CC S. mansoni or an active fragment, homologue or variant, fused to a  
 CC bacterial, phage or viral protein. The vaccine containing the fusion  
 CC protein has been found to induce a significantly increased antibody  
 CC response against schistosoma infections, compared to the use of S.  
 CC mansoni cercarial elastase in its native form. The present sequence  
 CC represents the S. mansoni cercarial elastase protein fragment encoded by  
 CC exon 2 of the cercarial elastase gene, used in the fusion protein of the  
 CC invention.

XX  
 SQ Sequence 106 AA;  
 Query Match 100.0%; Score 90; DB 21; Length 106;  
 Best Local Similarity 100.0%; Pred. No. 1.7e-07;  
 Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 VGYGRDNDNDPSPSRKN 16  
 DB 85 VGYGRDNDNDPSPSRKN 100  
 RESULT 3  
 ID AAW21484 standard; peptide; 11 AA.  
 AC AAW21484;  
 XX  
 DT 30-JUL-1997 (first entry)  
 XX  
 DE Schistosoma elastase precursor derived signal oligopeptide #4.  
 XX  
 KW Hydrophilic; signal oligopeptide; hydrophilicity maxima; vaccine; HIV;  
 KW competitive inhibitor; feedback regulator; synthesis; gastrin precursor;  
 KW charge; polarity; farnesyl synthetase; plasminogen activator inhibitor 1;  
 KW hydroxymethylglutaryl coenzyme A reductase; glucagon precursor; rheus;  
 KW gonadoliberin precursor; plasminogen activator inhibitor 2; prorenin;  
 KW Alzheimer amyloid A4; corticotropin releasing factor binding protein;  
 KW apolipoprotein E; herpes virus 1 glycoprotein B; HSV1; human; OMVVS;  
 KW herpes virus 2 glycoprotein B; HSV2; collagenase; apolipoprotein A;  
 KW Treponema pallidum membrane protein; TmPA; Islet amyloid polypeptide;  
 KW fibroblast MMP1; schistosoma elastase precursor; schistosomin;  
 KW hepatitis delta antigen; rev protein; HIV; VIHIV; angiotensinogen.  
 XX  
 OS Schistosoma sp.  
 XX  
 ZN W09519568-A1.  
 XX  
 ZD 20-JUL-1995.  
 XX  
 ZP 12-JAN-1995; 95WO-US00575.  
 XX  
 PR 14-JAN-1994; 94US-0182248.  
 XX  
 PA (RATH/) RATH M.  
 XX  
 PI Rath M;  
 XX  
 DR WPI; 1995-263953/34.  
 XX  
 PT Identifying signal oligopeptide(s) in protein sequence(s) - shown as  
 PT regions of max. hydrophilicity, used in modulating communication  
 PT between protein(s)  
 XX  
 Claim 5; Page 70; 88pp; English.  
 PS  
 CC The sequences given in AAW21201-560 represent hydrophilic signal oligo-  
 CC peptides. These signal oligopeptides are localised on the surface  
 CC of the protein and are represented by the hydrophilicity maxima of  
 CC the protein. These peptides are enriched in charged amino acids  
 CC arranged with neutral spacer amino acids. The specific signal  
 CC character of these oligopeptides is determined by a characteristic  
 CC combination of conformation and charge within the signal sequence.  
 CC These oligopeptides may be used as vaccines in the treatment of  
 CC human disease, as competitive inhibitors to prevent or reduce the  
 CC metabolic action or interaction of a selected protein by blocking  
 CC its specific signal sequences, or as therapeutic agents to function  
 CC as feedback regulators to reduce synthesis rate of a selected protein.  
 CC These peptides may be modified by omitting one or more amino acids at  
 CC the N- and/or C-terminal, by substituting one or more amino acids  
 CC without consideration of charge and polarity, by substituting one or  
 CC more amino acids with amino acid residues with similar charge and/or  
 CC polarity, by omitting one or more amino acids or a combination of these.





XX 06-SEP-2001; 2001WO-US27628.  
PF  
XX 06-SEP-2000; 2000US-230505P.  
PR 06-SEP-2000; 2000US-230514P.  
PR 06-SEP-2000; 2000US-230515P.  
PR 06-SEP-2000; 2000US-230517P.  
PR 06-SEP-2000; 2000US-230518P.  
PR 06-SEP-2000; 2000US-230519P.  
PR 06-SEP-2000; 2000US-230595P.  
PR 06-SEP-2000; 2000US-230597P.  
PR 06-SEP-2000; 2000US-230598P.  
PR 06-SEP-2000; 2000US-230599P.  
PR 06-SEP-2000; 2000US-230610P.  
PR 06-SEP-2000; 2000US-230659P.  
PR 06-SEP-2000; 2000US-230988P.  
PR 06-SEP-2000; 2000US-230989P.  
PR 07-SEP-2000; 2000US-230951P.  
PR 07-SEP-2000; 2000US-231163P.  
PR 07-SEP-2000; 2000US-231167P.  
XX  
PA (INCY-) INCYTE GENOMICS INC.  
XX  
PI Jackson S, Lincoln SE, Altus CM, Dufour GE, Chalup MS, Hillman JL;  
PI Jones AL, Yu JY, Wright RJ, Gietzen D, Liu TF, Yap PE, Dahl CR;  
PI Montiyama MG, Bradley DL, Rohatgi SD, Harris B, Roseberry AM;  
PI Gerstin EH, Peralta CH, David MH, Panzer SR, Flores V, Daffo A;  
PI Marwaha R, Chen AJ, Chang SC, Au AP, Iman RR;  
XX  
XX WPI; 2002-527544/56.  
DR N-PSDB; ABQ72512.  
XX  
XX Novel human disease detection and treatment polypeptide, useful in  
PT diagnosis, prevention or treatment of cell proliferative disorders e.g.  
PT arteriosclerosis, cirrhosis and an autoimmune/inflammatory disorder  
PT e.g. AIDS -  
XX  
XX Claim 14; Page 485-486; 618bp; English.  
PS  
XX The invention relates to an isolated human disease detection and  
CC treatment (MDDT) polypeptide (I) selected from a polypeptide having a  
CC sequence selected from 254 sequences (ABP51231-ABP51484) given in the  
CC specification, a naturally occurring polypeptide comprising a sequence  
CC having at least 90% identity to (I) or a biologically active or  
CC immunogenic fragment of (I). (I) is useful for screening a compound for  
CC effectiveness as an agonist or antagonist, for screening a compound that  
CC specifically binds (I) or modulates the activity of (I), and for  
CC preparing a polyclonal or monoclonal antibody by hybridoma technology.  
CC Nucleic acids (II) (ABQ72449-ABQ72700) encoding (I) are useful for  
CC screening a compound for effectiveness in altering expression of a target  
CC polynucleotide comprising. Oligonucleotides and antibodies are useful for  
CC detecting MDDT in a sample or for assessing toxicity of a test compound,  
CC in a diagnostic test for a condition or a disease associated with the  
CC expression of MDDT in a biological sample, for detecting (I) in a sample,  
CC and for purifying (I) from a sample. A composition comprising (I), an  
CC agonist or antagonist is useful for treating a disease or condition  
CC associated with decreased or increased expression of functional MDDT.  
CC (I) or (II) are useful for diagnosing, treating or preventing disorders  
CC associated with aberrant expression of MDDT, where the disorders are  
CC selected from a cell proliferative disorder such as arteriosclerosis,  
CC cirrhosis, hepatitis, psoriasis, and cancer and an  
CC autoimmune/inflammatory disorder such as AIDS, Addison's disease,  
CC allergy, anaemia, asthma, atherosclerosis, gout, multiple sclerosis or  
CC rheumatoid arthritis. (II) are useful for creating knockin humanised  
CC animals or transgenic animals to model human diseases. In somatic or  
CC germ-line gene therapy, to generate a transcript image of a tissue or cell  
CC type, for detecting differences in the chromosomal location due to  
CC translocation or inversion among normal, carrier or affected individuals  
CC and as hybridisation probes for mapping naturally occurring genomic  
CC sequences.  
XX  
XX WPI; 2002-383054/41.  
DR N-PSDB; ABK71715.  
XX  
XX An isolated polynucleotide useful in diagnostics and therapeutics -  
FT  
XX  
XX Sequence 757 AA;  
SQ

Query Match 52.2%; Score 47; DB 23; Length 757;  
Best Local Similarity 66.7%; Pred. No. 27;  
Matches 8; Conservative 2; Mismatches 2; Indels 0; Gaps 0;  
QY 5 RDDNDRPPSRKN 16  
Db 568 REDSDQPSMRN 579  
RESULT 6  
ABG60124  
ID ABG60124 standard; Protein; 762 AA.  
XX  
AC ABG60124;  
XX  
DT 30-JUL-2002 (first entry)  
XX  
DE Human DITHP polypeptide #182.  
XX  
KW Human; DITHP; diagnostic and therapeutic polypeptide; bone; testis; skin;  
KW cell proliferative disorder; cancer; tumour; autoimmune disorder; brain;  
KW inflammatory disorder; viral infection; bacterial infection; seizure;  
KW fungal infection; parasitic infections; developmental disorder; breast;  
KW endocrine disorder; metabolic disorder; neurological disorder; cervix;  
KW gastrointestinal disorder; transport disorder; gene therapy; kidney;  
KW adrenal gland; bone marrow; lung; ovary; pancreas; prostate; spleen;  
KW thymus.  
XX  
OS Homo sapiens.  
XX  
XX WO200220754-A2.  
XX  
XX 14-MAR-2002.  
XX  
XX 29-AUG-2001; 2001WO-US27127.  
XX  
XX 05-SEP-2000; 2000US-229747P.  
XX  
XX 05-SEP-2000; 2000US-229748P.  
XX  
XX 05-SEP-2000; 2000US-229749P.  
XX  
XX 05-SEP-2000; 2000US-229750P.  
XX  
XX 05-SEP-2000; 2000US-229751P.  
XX  
XX 05-SEP-2000; 2000US-230583P.  
XX  
XX 06-SEP-2000; 2000US-230505P.  
XX  
XX 06-SEP-2000; 2000US-230514P.  
XX  
XX 06-SEP-2000; 2000US-230515P.  
XX  
XX 06-SEP-2000; 2000US-230517P.  
XX  
XX 06-SEP-2000; 2000US-230518P.  
XX  
XX 06-SEP-2000; 2000US-230519P.  
XX  
XX 06-SEP-2000; 2000US-230595P.  
XX  
XX 06-SEP-2000; 2000US-230597P.  
XX  
XX 06-SEP-2000; 2000US-230598P.  
XX  
XX 06-SEP-2000; 2000US-230599P.  
XX  
XX 06-SEP-2000; 2000US-230610P.  
XX  
XX 06-SEP-2000; 2000US-230659P.  
XX  
XX 06-SEP-2000; 2000US-230988P.  
XX  
XX 07-SEP-2000; 2000US-230951P.  
XX  
XX 07-SEP-2000; 2000US-231163P.  
XX  
XX 07-SEP-2000; 2000US-231167P.  
XX  
XX (INCY-) INCYTE GENOMICS INC.  
XX  
XX Stuart J, Lincoln SE, Altus CM, Dufour GE, Chalup MS, Hillman JL;  
XX Jones AL, Yu JY, Wright RJ, Gietzen D, Liu TF, Yap PE, Dahl CR;  
XX Montiyama MG, Bradley DL, Rohatgi SD, Harris B, Roseberry AM;  
XX Gerstin EH, Peralta CH, David MH, Panzer SR, Flores V, Daffo A;  
XX Marwaha R, Chen AJ, Chang SC, Au AP, Iman RR;  
XX  
XX WPI; 2002-383054/41.  
DR N-PSDB; ABK71715.  
XX  
XX An isolated polynucleotide useful in diagnostics and therapeutics -  
FT  
XX  
XX Claim 29; Page 637-639; 686bp; English.  
FS

XX The invention relates to human diagnostic and therapeutic (dthp)  
 CC polynucleotides and their associated polypeptides (DTHP polypeptides).  
 CC The sequences of the invention are used in the treatment and diagnosis of  
 CC cell proliferative disorders (e.g. atherosclerosis, cirrhosis), cancers  
 CC (e.g. tumours of the adrenal gland, bone, bone marrow, brain, breast,  
 CC cervix, kidney, lung, ovary, pancreas, prostate, skin, spleen, testis or  
 CC thymus), autoimmune/inflammatory disorders (e.g. asthma, bronchitis,  
 CC psoriasis, osteoporosis), viral infections, bacterial infections, fungal  
 CC infections, parasitic infections, developmental disorders (e.g. anaemia,  
 CC epilepsy), seizure disorders (e.g. cerebral palsy, spina bifida),  
 CC endocrine disorders (e.g. thrombosis, aneurysm), metabolic disorders  
 CC (e.g. obesity, diabetes), neurological disorders (e.g. stroke,  
 CC amyotrophic lateral sclerosis, multiple sclerosis), gastrointestinal  
 CC disorders (e.g. ulcerative colitis, lysinuria) and transport disorders  
 CC (e.g. myotonic dystrophy, catatonias, peripheral neuropathy). Sequences  
 CC ABG59943-ABG60220 represent human DTHP polypeptides of the invention.  
 XX  
 SQ Sequence 762 AA;  
 Query Match 52.2%; Score 47; DB 23; Length 762;  
 Best Local Similarity 66.7%; Pred. No. 27;  
 Matches 8; Conservative 2; Mismatches 2; Indels 0; Gaps 0;  
 QY 5 RDDNDRDPSSRN 16  
 Db 568 REDDDQDPSMKN 579  
 RESULT 7  
 ABP47797  
 ID ABP47797 standard; Protein; 1429 AA.  
 XX  
 AC ABP47797;  
 DT 21-AUG-2002 (first entry)  
 XX  
 DE Yeast DNA topoisomerase.  
 XX  
 KM Enzyme inhibitor; molecular structure; antiviral; anti-HIV;  
 KM human immunodeficiency virus.  
 XX  
 OS Saccharomyces cerevisiae.  
 XX  
 PN WO200244112-A1.  
 XX  
 PD 06-JUN-2002.  
 XX  
 PF 30-NOV-2000; 2000WO-JP08467.  
 XX  
 PR 30-NOV-2000; 2000WO-JP08467.  
 XX  
 PA (TOSU-) TOYO SUISAN KAISHA LTD.  
 XX  
 PI Mizushima Y, Sakaguchi K, Sugawara F;  
 XX  
 DR WPI; 2002-471820/50.  
 XX  
 PT Comparison of enzyme-inhibitor interaction of a reference and target  
 PT enzyme using HMQC NMR for design of molecular structure of inhibitors  
 PT to the target enzyme -  
 XX  
 PS Example 2; Page 102-108; 140pp; Japanese.  
 XX  
 CC The invention relates to a novel method for determining the molecular  
 CC structure of an inhibitor to a target enzyme. The method of the invention  
 CC has antiviral activity. The method is useful for identification of  
 CC inhibitors to a target enzyme, in particular for use as anti-HIV (human  
 CC immunodeficiency virus) agents, without the need for X-ray  
 CC crystallography of the enzyme. The sequence represents a protein related  
 CC to the invention.  
 XX  
 SQ Sequence 1429 AA;

Query Match 51.1%; Score 46; DB 23; Length 1429;  
 Best Local Similarity 72.7%; Pred. No. 82;  
 Matches 8; Conservative 1; Mismatches 2; Indels 0; Gaps 0;  
 QY 6 DDNDRDPSSRN 16  
 Db 1199 DDEDYDPSKKN 1209  
 RESULT 8  
 AAB10642  
 ID AAB10642 standard; Protein; 149 AA.  
 XX  
 AC AAB10642;  
 XX  
 DT 19-JAN-2001 (first entry)  
 XX  
 DE Human VEGF-X PDGF-like domain protein.  
 XX  
 KM VEGF-X; vascular endothelial growth factor; human; vulnery; cytostatic;  
 KM antirheumatic; antiarthritic; antiporiatic; antidiabetic; treatment;  
 KM angiogenesis regulator; vascularization regulator; cancer; psoriasis;  
 KM rheumatoid arthritis; diabetic retinopathy; blood vessel; organ repair;  
 KM tissue regeneration; tissue repair; wound; dermal ulcer; pressure sore;  
 KM venous sore; diabetic ulcer; burns; skin graft growth.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO200037641-A2.  
 XX  
 PD 29-JUN-2000.  
 XX  
 PF 21-DEC-1999; 99NO-US30503.  
 XX  
 PR 22-DEC-1998; 98GB-0028377.  
 PR 18-MAR-1999; 99US-0124967.  
 PR 08-NOV-1999; 99US-0164131.  
 XX  
 PA (JUNC ) JANSSEN PHARM NV.  
 XX  
 PI Gordon RD, Sprengel J, Yon JR, Dijkmans JH, Gosiowska A;  
 PI Dhanaraj SN, Xu J;  
 XX  
 DR WPI; 2000-442669/38.  
 DR N-PSDB; AAA71986.  
 XX  
 PT New vascular endothelial growth factor protein, useful for treating or  
 PT preventing diseases associated with inappropriate angiogenesis activity  
 PT such as cancer, rheumatoid arthritis, psoriasis and wounds -  
 XX  
 PS Disclosure; Fig 24; 127pp; English.  
 XX  
 CC This invention describes a novel vascular endothelial growth factor-X  
 CC (VEGF-X) protein (ia) and its encoding polynucleotide (Iia) which has  
 CC vulnery, cytostatic, antirheumatic, antiarthritic, antiporiatic and  
 CC antidiabetic activity and acts as an angiogenesis and vascularization  
 CC regulator. An antisense molecule of the invention is useful for treating  
 CC or preventing cancer, rheumatoid arthritis, psoriasis and diabetic  
 CC retinopathy by inhibiting angiogenic activity or inappropriate  
 CC vascularization including formation and proliferation of new blood  
 CC vessels, growth and development of tissues, tissue regeneration and organ  
 CC and tissue repair in a subject. The products of the invention are useful  
 CC for preparing medicaments for treating wounds such as dermal ulcers,  
 CC pressure sores, venous sores, diabetic ulcers and burns and to promote  
 CC skin graft growth, tissue repair, proliferation of new blood vessels,  
 CC tissue regeneration and organ repair by promoting angiogenic activity or  
 CC vascularization. This sequence represents a human VEGF-X protein  
 CC PDGF-like domain which can be expressed in E. coli systems and which is  
 CC described in the method of the invention.  
 XX  
 SQ Sequence 149 AA;

Query Match 50.0%; Score 45; DB 21; Length 149;  
Best Local Similarity 63.6%; Pred. No. 9.5;  
Matches 7; Conservative 3; Mismatches 1; Indels 0; Gaps 0;  
QY 6 DDNDPDRKRN 16  
||:|:|:|:  
Db 27 DDDKDPGRKS 37

RESULT 9  
AAc46818  
ID AAc46818 standard; Protein; 717 AA.  
XX  
AC AAc46818;  
XX  
DT 18-OCT-2000 (first entry)  
XX  
DE Arabidopsis thaliana protein fragment SEQ ID NO: 58941.  
XX  
KW Protein identification; signal transduction pathway; metabolic pathway;  
hybridisation assay; genetic mapping; gene expression control; promoter;  
termination sequence.  
XX  
OS Arabidopsis thaliana.  
XX  
PN EP1033405-A2.  
XX  
PD 06-SEP-2000.  
XX  
PF 25-FEB-2000; 2000EP-0301439.  
XX  
PR 25-FEB-1999; 99US-0121825.  
PR 05-MAR-1999; 99US-0123180.  
PR 09-MAR-1999; 99US-0123548.  
PR 23-MAR-1999; 99US-0125788.  
PR 23-MAR-1999; 99US-0126264.  
PR 29-MAR-1999; 99US-0126785.  
PR 01-APR-1999; 99US-0127462.  
PR 06-APR-1999; 99US-0128234.  
PR 08-APR-1999; 99US-0128714.  
PR 16-APR-1999; 99US-0128845.  
PR 19-APR-1999; 99US-0130077.  
PR 21-APR-1999; 99US-0130449.  
PR 23-APR-1999; 99US-0130510.  
PR 23-APR-1999; 99US-0130891.  
PR 28-APR-1999; 99US-0131449.  
PR 30-APR-1999; 99US-0132048.  
PR 30-APR-1999; 99US-0132407.  
PR 04-MAY-1999; 99US-0132484.  
PR 05-MAY-1999; 99US-0132485.  
PR 06-MAY-1999; 99US-0132486.  
PR 06-MAY-1999; 99US-0132487.  
PR 07-MAY-1999; 99US-0132863.  
PR 11-MAY-1999; 99US-0132856.  
PR 14-MAY-1999; 99US-0134218.  
PR 14-MAY-1999; 99US-0134219.  
PR 14-MAY-1999; 99US-0134221.  
PR 14-MAY-1999; 99US-0134370.  
PR 18-MAY-1999; 99US-0134768.  
PR 19-MAY-1999; 99US-0134941.  
PR 20-MAY-1999; 99US-0135124.  
PR 21-MAY-1999; 99US-0135353.  
PR 24-MAY-1999; 99US-0135629.  
PR 25-MAY-1999; 99US-0136021.  
PR 27-MAY-1999; 99US-0136392.  
PR 28-MAY-1999; 99US-0136782.  
PR 01-JUN-1999; 99US-0137222.  
PR 03-JUN-1999; 99US-0137528.  
PR 04-JUN-1999; 99US-0137502.  
PR 07-JUN-1999; 99US-0137724.  
PR 08-JUN-1999; 99US-0138094.  
PR 10-JUN-1999; 99US-0138540.  
PR 10-JUN-1999; 99US-0138847.

PR 14-JUN-1999; 99US-0139119.  
PR 16-JUN-1999; 99US-0139452.  
PR 16-JUN-1999; 99US-0139453.  
PR 17-JUN-1999; 99US-0139492.  
PR 18-JUN-1999; 99US-0139454.  
PR 18-JUN-1999; 99US-0139455.  
PR 18-JUN-1999; 99US-0139456.  
PR 18-JUN-1999; 99US-0139457.  
PR 18-JUN-1999; 99US-0139458.  
PR 18-JUN-1999; 99US-0139459.  
PR 18-JUN-1999; 99US-0139460.  
PR 18-JUN-1999; 99US-0139461.  
PR 18-JUN-1999; 99US-0139462.  
PR 18-JUN-1999; 99US-0139463.  
PR 18-JUN-1999; 99US-0139750.  
PR 18-JUN-1999; 99US-0139763.  
PR 21-JUN-1999; 99US-0139817.  
PR 22-JUN-1999; 99US-0139889.  
PR 23-JUN-1999; 99US-0140353.  
PR 23-JUN-1999; 99US-0140354.  
PR 24-JUN-1999; 99US-0140695.  
PR 28-JUN-1999; 99US-0140823.  
PR 29-JUN-1999; 99US-0140991.  
PR 30-JUN-1999; 99US-0141287.  
PR 01-JUL-1999; 99US-0141842.  
PR 01-JUL-1999; 99US-0142154.  
PR 02-JUL-1999; 99US-0142055.  
PR 06-JUL-1999; 99US-0142390.  
PR 08-JUL-1999; 99US-0142803.  
PR 09-JUL-1999; 99US-0142920.  
PR 12-JUL-1999; 99US-0142977.  
PR 13-JUL-1999; 99US-0143542.  
PR 14-JUL-1999; 99US-0143624.  
PR 15-JUL-1999; 99US-0144005.  
PR 16-JUL-1999; 99US-0144085.  
PR 16-JUL-1999; 99US-0144086.  
PR 19-JUL-1999; 99US-0144325.  
PR 19-JUL-1999; 99US-0144331.  
PR 19-JUL-1999; 99US-0144332.  
PR 19-JUL-1999; 99US-0144333.  
PR 19-JUL-1999; 99US-0144334.  
PR 19-JUL-1999; 99US-0144335.  
PR 20-JUL-1999; 99US-0144352.  
PR 20-JUL-1999; 99US-0144632.  
PR 20-JUL-1999; 99US-0144884.  
PR 21-JUL-1999; 99US-0144814.  
PR 21-JUL-1999; 99US-0145086.  
PR 21-JUL-1999; 99US-0145088.  
PR 22-JUL-1999; 99US-0145085.  
PR 22-JUL-1999; 99US-0145087.  
PR 22-JUL-1999; 99US-0145089.  
PR 22-JUL-1999; 99US-0145192.  
PR 23-JUL-1999; 99US-0145145.  
PR 23-JUL-1999; 99US-0145218.  
PR 23-JUL-1999; 99US-0145224.  
PR 26-JUL-1999; 99US-0145276.  
PR 27-JUL-1999; 99US-0145913.  
PR 27-JUL-1999; 99US-0145918.  
PR 27-JUL-1999; 99US-0145919.  
PR 28-JUL-1999; 99US-0145951.  
PR 28-JUL-1999; 99US-0146386.  
PR 02-AUG-1999; 99US-0146388.  
PR 02-AUG-1999; 99US-0146389.  
PR 03-AUG-1999; 99US-0147038.  
PR 04-AUG-1999; 99US-0147204.  
PR 04-AUG-1999; 99US-0147302.  
PR 05-AUG-1999; 99US-0147192.  
PR 05-AUG-1999; 99US-0147260.  
PR 06-AUG-1999; 99US-0147303.  
PR 06-AUG-1999; 99US-0147416.  
PR 09-AUG-1999; 99US-0147493.  
PR 09-AUG-1999; 99US-0147935.  
PR 10-AUG-1999; 99US-0148171.

```
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148365.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 13-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158028.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159328.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.
PR 29-OCT-1999; 99US-0162142.
```

Query Match 50.0%; Score 45; DB 21; Length 717;  
Best Local Similarity 72.7%; Pred. No. 56;  
Matches 8; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 4 GRDNDNRDPSR 14  
DB 124 GLDDRRDRDPSR 134

RESULT 10  
ID ABG15848  
ABG15848 standard; Protein: 1224 AA.  
XX AC ABG15848;  
XX DT 18-FEB-2002 (first entry)  
XX DE Novel human diagnostic protein #15839.  
XX KW Human; chromosome mapping; gene mapping; gene therapy; forensic;  
XX KW food supplement; medical imaging; diagnostic; genetic disorder.  
XX OS Homo sapiens.  
XX PN WO200175067-A2.  
XX PD 11-OCT-2001.  
XX PE 30-MAR-2001; 2001WO-US08631.  
XX PR 31-MAR-2000; 2000US-0540217.  
XX PR 23-AUG-2000; 2000US-0649167.  
XX PA (HYSE-) HYSEQ INC.  
XX PI Dmanac RT, Liu C, Tang YT;  
XX DR WPI; 2001-639362/73.  
XX DR N-PSDB; AAS80035.  
XX PT New isolated polynucleotide and encoded polypeptides, useful in  
PT diagnostics, forensics, gene mapping, identification of mutations  
PT responsible for genetic disorders or other traits and to assess  
PT biodiversity -  
XX PS Claim 20; SEQ ID No 46207; 103pp; English.  
XX CC The invention relates to isolated polynucleotide (I) and  
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,  
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome  
CC and gene mapping, and in recombinant production of (II). The  
CC polynucleotides are also used in diagnostics as expressed sequence tags  
CC for identifying expressed genes. (I) is useful in gene therapy techniques  
CC to restore normal activity of (II) or to treat disease states involving  
CC (II). (II) is useful for generating antibodies against it, detecting or  
CC quantitating a polypeptide in tissue, as molecular weight markers and as  
CC a food supplement. (II) and its binding partners are useful in medical  
CC imaging of sites expressing (II). (I) and (II) are useful for treating  
CC disorders involving aberrant protein expression or biological activity.  
CC The polypeptide and polynucleotide sequences have applications in  
CC diagnostics, forensics, gene mapping, identification of mutations  
CC responsible for genetic disorders or other traits to assess biodiversity  
CC and to produce other types of data and products dependent on DNA and  
CC amino acid sequences. ABG00010-ABG30377 represent novel human  
CC diagnostic amino acid sequences of the invention.  
CC Note: The sequence data for this patent did not appear in the printed  
CC specification, but was obtained in electronic format directly from WIPO  
XX at ftp.wipo.int/pub/published\_pct\_sequences.  
SQ Sequence 1224 AA;

Query Match 50.0%; Score 45; DB 22; Length 1224;  
Best Local Similarity 66.7%; Pred. No. 1e+02;  
Matches 8; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 2 GYGRDNDNRDPS 13  
DB 115 GYGFRDNDNRDPS 126

RESULT 11  
AAU09996 ID AAU09996 standard; Protein; 614 AA.  
XX  
AC AAU09996;  
XX  
DT 21-MAY-2002 (first entry)  
XX  
DE Geniculosporium haloperoxidase protein.  
XX  
KM Haloperoxidase; antimicrobial; antibacterial; detergent; paint;  
XX ophthalmological; oxidation; halide.  
XX  
OS Geniculosporium sp.  
XX  
PN WO200179460-A2.  
XX  
PD 25-OCT-2001.  
XX  
PF 10-APR-2001; 2001WO-DK00242.  
XX  
PR 14-APR-2000; 2000DK-0000626.  
XX  
PA (NOVO ) NOVOZYMES AS.  
XX  
PI Danielson S, Schneider P;  
XX  
DR WPI; 2002-011127/01.  
DR N-PSDB; AAI6215.  
XX  
PT New isolated polypeptide with haloperoxidase activity useful as, e.g.,  
XX a microbicide or detergent additive -  
XX  
PS Claim 3; Page 48-49; 49pp; English.  
XX  
CC This sequence represents a geniculosporium haloperoxidase protein of  
CC the invention. The invention also comprises the sequence of the  
CC geniculosporium haloperoxidase gene. Haloperoxidases may have  
CC antimicrobial, antibacterial and ophthalmological activities. The  
CC haloperoxidase protein is useful to catalyse the oxidation of a halide  
CC ion in the presence of peroxide (H<sub>2</sub>O<sub>2</sub>) and may be used in DNA  
CC recombination processes or to halogenate compounds. They may be added  
CC to detergent compositions or for killing or inhibiting growth of  
CC microbial cells. They may be also be used as preservatives or  
CC disinfectants in, e.g., paints, personal care products or contact lens  
CC cleaning products or in cleaning or bleaching compositions. The  
CC nucleotide and protein sequences given in the specification may also be  
CC used to design nucleic acid probes to identify and clone DNA encoding  
CC polypeptides with haloperoxidase activity.  
XX  
SQ Sequence 614 AA;  
Query Match 48.9%; Score 44; DB 23; Length 614;  
Best Local Similarity 80.0%; Pred. No. 69;  
Matches 8; Conservative 1; Mismatches 1; Indels 0; Gaps 0;  
QY 5 RDDNDRDPSR 14  
DB 366 RDDNFRDPMR 375  
RESULT 12  
AAG78515 ID AAG78515 standard; Protein; 614 AA.  
XX  
AC AAG78515;  
XX  
DT 29-JAN-2002 (first entry)  
XX  
DE Haloperoxidase amino acid sequence.  
XX  
KM Haloperoxidase; Oxidation; Antimicrobial; Antibacterial; Ophthalmological;  
KM Chloroperoxidase; Bromoperoxidase; Iodoperoxidase; Halogen; Detergent;  
XX

KM Preservative; Disinfectant; Paint; Contact lens cleaning; Bleaching.  
XX  
OS Geniculosporium sp.  
XX  
PN WO200179463-A2.  
XX  
PD 25-OCT-2001.  
XX  
PF 10-APR-2001; 2001WO-DK00246.  
XX  
PR 14-APR-2000; 2000DK-0000626.  
XX  
PA (NOVO ) NOVOZYMES AS.  
XX (MAXY-) MAXYGEN INC.  
XX  
PI Danielson S, Schneider P;  
XX  
DR WPI; 2002-011128/01.  
DR N-PSDB; AAI64222.  
XX  
PT New isolated nucleic acid sequence comprising a sequence encoding a  
XX polypeptide with haloperoxidase activity useful as a microbicide -  
XX  
PS Claim 1; Page 48-50; 50pp; English.  
XX  
CC The invention related to an isolated nucleic acid sequence comprising  
CC a sequence encoding a polypeptide with haloperoxidase activity.  
CC Haloperoxidase activity can be described as: antimicrobial,  
CC antibacterial and ophthalmological. The antibacterial activity of  
CC Geniculosporium sp. haloperoxidase was tested against *Escherichia coli*  
CC DSM 1576 using bromide as enhancing agent. A combination of 6 mM KBr,  
CC 1 mg/L enzyme and 1 mM H<sub>2</sub>O<sub>2</sub> gave complete kill of the bacteria.  
CC The nucleic acids encode polypeptides with haloperoxidase activity  
CC (i.e., they catalyse the oxidation of a halide ion in the presence of  
CC H<sub>2</sub>O<sub>2</sub>) and may be used in DNA recombination processes. They may be added  
CC to detergent compositions or for killing or inhibiting growth of  
CC microbial cells. They may be also be used as preservatives or  
CC disinfectants in, e.g., paints, personal care products or contact lens  
CC cleaning products or in cleaning or bleaching compositions. They may also  
CC be used in enzymatic bleaching applications, such as pulp bleaching and  
CC stain bleaching. Polypeptides and polynucleotides of the invention may  
CC also be used to design nucleic acid probes to identify and clone DNA  
CC encoding polypeptides with haloperoxidase activity. The current sequence  
CC represents a haloperoxidase amino acid sequence.  
XX  
SQ Sequence 614 AA;  
Query Match 48.9%; Score 44; DB 23; Length 614;  
Best Local Similarity 80.0%; Pred. No. 69;  
Matches 8; Conservative 1; Mismatches 1; Indels 0; Gaps 0;  
QY 5 RDDNDRDPSR 14  
DB 366 RDDNFRDPMR 375  
RESULT 13  
AAU43208 ID AAU43208 standard; Protein; 84 AA.  
XX  
AC AAU43208;  
XX  
DT 27-FEB-2002 (first entry)  
XX  
DE Propionibacterium acnes immunogenic protein #4104.  
XX  
KM SAPHO syndrome; synovitis; acne; pustulosis; hypertosis; osteomyelitis;  
KM uveitis; endophthalmitis; bone; joint; central nervous system; ELISA;  
KM inflammatory lesion; acne vulgaris; enzyme linked immunosorbent assay;  
KM dermatological; osteoporotic; neuroprotectant.  
XX  
OS Propionibacterium acnes.  
XX

PN WO200181581-A2.  
 XX  
 PD 01-NOV-2001.  
 XX  
 PF 20-APR-2001; 2001WO-US12865.  
 XX  
 PR 21-APR-2000; 2000US-199047P.  
 PR 02-JUN-2000; 2000US-208841P.  
 PR 07-JUL-2000; 2000US-216747P.  
 XX  
 PA (CORI-) CORIXA CORP.  
 XX  
 PI Sheiky YM, Persing DH, Mitcham JL, Wang SS, Bhavia A;  
 PI L.malsomneuve J, Zhang Y, Jen S, Carter D;  
 XX  
 DR WPI; 2001-616774/71.  
 DR N-PSDB; AAS59520.  
 XX  
 PT Propionibacterium acnes polypeptides and nucleic acids useful for  
 PT vaccinating against and diagnosing infections, especially useful for  
 PT treating acne vulgaris -  
 XX  
 XX Example 1; SEQ ID NO 4403; 1069bp; English.  
 PS  
 CC Sequences AAU39105-AAU68017 represent Propionibacterium acnes immunogenic  
 CC polypeptides. The proteins and their associated DNA sequences are used in  
 CC the treatment, prevention and diagnosis of medical conditions caused by  
 CC P. acnes. The disorders include SAPHO syndrome (synovitis, acne,  
 CC pustulosis, hyperostosis and osteomyelitis), uveitis and endophthalmitis.  
 CC P. acnes is also involved in infections of bone, joints and the central  
 CC nervous system, however it is particularly involved in the inflammatory  
 CC lesions associated with acne vulgaris. A method for detecting the  
 CC presence or absence of P. acnes in a patient comprises contacting a  
 CC sample with a binding agent that binds to the proteins of the invention  
 CC and determining the amount of bound protein in the sample. The  
 CC polypeptides may be used as antigens in the production of antibodies  
 CC specific for P. acnes proteins. These antibodies can be used to  
 CC downregulate expression and activity of P. acnes polypeptides and  
 CC therefore treat P. acnes infections. The antibodies may also be used as  
 CC diagnostic agents for determining P. acnes presence, for example, by  
 CC enzyme linked immunosorbent assay (ELISA).  
 CC Note: The sequence data for this patent did not form part of the printed  
 CC specification, but was obtained in electronic format directly from WIPO  
 CC at fcp.wipo.int/pub/published\_pct\_sequences.  
 CC  
 XX  
 SQ Sequence 84 AA;  
 Query Match 47.8%; Score 43; DB 22; Length 84;  
 Best Local Similarity 66.7%; Pred. No. 11;  
 Matches 8; Conservative 1; Mismatches 3; Indels 0; Gaps 0;  
 QY 4 GRDNDRDPSRK 15  
 |||  
 Db 9 GRGDYDRDPVRR 20  
 |||  
 RESULT 14  
 AAM40752  
 ID AAM40752 standard; Protein; 229 AA.  
 XX  
 AC AAM40752;  
 XX  
 DT 22-OCT-2001 (first entry)  
 XX  
 DE Human polypeptide SEQ ID NO 5683.  
 XX  
 KM Human; noctropic; immunosuppressant; cytostatic; gene therapy; cancer;  
 KM peripheral nervous system; neuropathy; central nervous system; CNS;  
 KM Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;  
 KM amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;  
 KM chemokinetic; thrombolytic; drug screening; arthritis; inflammation;  
 KM leukemia.  
 XX

OS Homo sapiens.  
 XX  
 PN WO200153312-A1.  
 XX  
 PD 26-JUL-2001.  
 XX  
 PF 26-DEC-2000; 2000WO-US34263.  
 XX  
 PR 21-JAN-2000; 2000US-0488725.  
 PR 25-APR-2000; 2000US-0552317.  
 PR 09-JUL-2000; 2000US-0598042.  
 PR 19-JUL-2000; 2000US-0620312.  
 PR 03-AUG-2000; 2000US-0653450.  
 PR 14-SEP-2000; 2000US-0662191.  
 PR 19-OCT-2000; 2000US-0693036.  
 PR 29-NOV-2000; 2000US-0727344.  
 XX  
 PA (HYSE-) HYSEQ INC.  
 XX  
 PI Tang YT, Liao C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;  
 PI Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J;  
 PI Zhao QA, Zhou P, Goodrich R, Drmanac RT;  
 XX  
 DR WPI; 2001-442253/47.  
 DR N-PSDB; AAI59908.  
 XX  
 PT Novel nucleic acids and polypeptides, useful for treating disorders  
 PT such as central nervous system injuries -  
 XX  
 XX Example 2; SEQ ID NO 5683; 10078bp; English.  
 PS  
 CC The invention relates to human nucleic acids (AAI57798-AAI61369) and  
 CC the encoded polypeptides (AAM38642-AAM42213) with nootropic.  
 CC immunosuppressant and cytostatic activity. The polynucleotides are useful  
 CC in gene therapy. A composition containing a polypeptide or polynucleotide  
 CC of the invention may be used to treat diseases of the peripheral nervous  
 CC system, such as peripheral nervous injuries, peripheral neuropathy and  
 CC localised neuropathies and central nervous system diseases, such as  
 CC Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic  
 CC lateral sclerosis, and Shy-Drager Syndrome. Other uses include the  
 CC utilisation of the activities such as: Immune system suppression,  
 CC Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic  
 CC and thrombolytic activity, cancer diagnosis and therapy, drug screening,  
 CC assays for receptor activity, arthritis and inflammation, leukaemias and  
 CC C.N.S disorders  
 CC Note: The sequence data for this patent did not form part of the printed  
 CC specification.  
 CC  
 XX  
 SQ Sequence 229 AA;  
 Query Match 47.8%; Score 43; DB 22; Length 229;  
 Best Local Similarity 66.7%; Pred. No. 33;  
 Matches 8; Conservative 1; Mismatches 3; Indels 0; Gaps 0;  
 QY 4 GRDNDRDPSRK 15  
 |||  
 Db 188 GRGTHDRDPSEK 199  
 |||  
 RESULT 15  
 AAG82400  
 ID AAG82400 standard; Protein; 241 AA.  
 XX  
 AC AAG82400;  
 XX  
 DT 03-SEP-2001 (first entry)  
 XX  
 DE S. epidermidis open reading frame protein sequence SEQ ID NO:1894.  
 XX  
 KM Staphylococcus epidermidis SRI strain; infection; diagnosis;  
 KM vaccination; endocarditis.  
 KM Staphylococcus epidermidis.  
 OS

XX WO200134809-A2.  
PN  
PD 17-MAY-2001.  
XX  
PF 09-NOV-2000; 2000WO-US30782.  
XX  
PR 09-NOV-1999; 99US-0164258.  
XX  
PA (GLAXO ) GLAXO GROUP LTD.  
XX  
PI Kimmerly WJ;  
XX  
DR WPI; 2001-316495/33.  
DR N-PSDB; AAH53250.  
XX  
PT Nucleic acids encoding polypeptides from Staphylococcus epidermidis,  
PT useful for vaccinating against infections, e.g. endocarditis -  
XX  
PS Claim 18; Page 517; 2188pp; English.  
XX  
CC AAH52304 to AAH53970 represent nucleic acids (I) encoding polypeptides  
CC (II), given in AAG81454 to AAG83120, from Staphylococcus epidermidis.  
CC (I) and (II) can have antibacterial activity and therefore can be used  
CC in vaccination. The nucleic acids (I) may be used to produce the  
CC S. epidermidis polypeptides (II) via the production of vectors  
CC containing them which are used to produce hosts cells which express the  
CC polypeptides. The polypeptides (II) (and/or nucleic acids) may then be  
CC used to vaccinate subjects and to raise antibodies against the bacteria.  
CC The polypeptides may also be used to assay for other inhibitors of their  
CC activity and therefore identify compounds that may be used for the  
CC treatment of S. epidermidis infections, e.g. endocarditis. AAH53971 to  
CC AAH55090 represent specifically claimed S. epidermidis genomic DNA  
CC polynucleotide sequences from the present invention. AAH55091 to  
CC AAH55098 represent oligonucleotide sequences and primers which are used  
CC in the exemplification of the present invention.  
CC N.B. The present invention specifically claims all the polynucleotide  
CC sequences given in the sequence listing of the present specification,  
CC however the sequence listing only goes up to SEQ ID NO:4454 so even  
CC though sequences are given in the disclosure for SEQ ID NO:4465 to 4472,  
CC no sequences are present for SEQ ID NO:4455 to 4464.  
XX  
SQ Sequence 241 AA;  
Query Match 47.8%; Score 43; DB 22; Length 241;  
Best Local Similarity 77.8%; Pred. No. 35;  
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;  
QY 2 GYGRDDNDR 10  
Db 166 GYNNNDNR 174  
RESULT 16  
ABP39919  
ID ABP39919 standard; Protein; 258 AA.  
XX  
AC ABP39919;  
XX  
DT 24-JUL-2002 (first entry)  
XX  
DE Staphylococcus epidermidis ORF amino acid sequence SEQ ID NO:4764.  
XX  
KM Staphylococcus epidermidis; open reading frame; ORF; bacterial infection;  
XX antibacterial; gene therapy.  
XX  
OS Staphylococcus epidermidis.  
XX  
PN US6380370-B1.  
XX  
PD 30-APR-2002.  
XX  
PF 13-AUG-1998; 98US-0134001.

XX 14-AUG-1997; 97US-055779P.  
PR 08-NOV-1997; 97US-064964P.  
XX  
PA (GENO-) GENOME THERAPEUTICS CORP.  
XX  
PI Doucette-Stamm LA, Bush D;  
XX  
DR WPI; 2002-381255/41.  
DR N-PSDB; ABN92464.  
XX  
PT Novel isolated nucleic acid encoding a Staphylococcus epidermidis  
PT polypeptide, useful for diagnosing and treating bacterial infections -  
XX  
PS Disclosure; SEQ ID 4764; 267pp; English.  
XX  
CC ABN90538 to ABN93374 represent Staphylococcus epidermidis open reading  
CC frame (ORF) nucleic acid sequences which encode the amino acid sequences  
CC given in ABP35124 to ABP37960. The S. epidermidis sequences have  
CC antibacterial activity and can be used in gene therapy. The sequences  
CC can also be used in the diagnosis and treatment of bacterial infections,  
CC particularly S. epidermidis infections. The sequences can be used to  
CC screen for compounds able to interfere with the S. epidermidis life  
CC cycle or inhibit S. epidermidis infection.  
CC N.B. The sequence data for this patent did not form part of the printed  
CC specification, but was obtained in electronic format directly from the  
CC USPTO web site.  
XX  
SQ Sequence 258 AA;  
Query Match 47.8%; Score 43; DB 23; Length 258;  
Best Local Similarity 77.8%; Pred. No. 38;  
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;  
QY 2 GYGRDDNDR 10  
Db 183 GYNNNDNR 191  
RESULT 17  
ABG10604  
ID ABG10604 standard; Protein; 297 AA.  
XX  
AC ABG10604;  
XX  
DT 13-FEB-2002 (first entry)  
XX  
DE Novel human diagnostic protein #10595.  
XX  
KM Human; chromosome mapping; gene mapping; gene therapy; forensic;  
XX food supplement; medical imaging; diagnostic; genetic disorder.  
XX  
OS Homo sapiens.  
XX  
PN WO200175067-A2.  
XX  
PD 11-OCT-2001.  
XX  
PF 30-MAR-2001; 2001WO-US08631.  
XX  
PR 31-MAR-2000; 2000US-0540217.  
PR 23-AUG-2000; 2000US-0649167.  
XX  
PA (HYSE-) HYSEQ INC.  
XX  
PI Drmanac RT, Liu C, Tang YT;  
XX  
DR WPI; 2001-639362/73.  
DR N-PSDB; AAS74791.  
XX  
PT New isolated polynucleotide and encoded polypeptides, useful in  
PT diagnostics, forensics, gene mapping, identification of mutations  
PT responsible for genetic disorders or other traits and to assess



PT biodiversity -  
 XX  
 PS Claim 20; SEQ ID NO 40963; 103bp; English.  
 XX  
 CC The invention relates to isolated polynucleotide (I) and  
 CC polypeptide (II) sequences. (I) is useful as hybridization probes,  
 CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome  
 CC and gene mapping, and in recombinant production of (II). The  
 CC polynucleotides are also used in diagnostics as expressed sequence tags  
 CC for identifying expressed genes. (I) is useful in gene therapy techniques  
 CC to restore normal activity of (II) or to treat disease states involving  
 CC (II). (II) is useful for generating antibodies against it, detecting or  
 CC quantitating a polypeptide in tissue, as molecular weight markers and as  
 CC a food supplement. (II) and its binding partners are useful in medical  
 CC imaging of sites expressing (II). (I) and (II) are useful for treating  
 CC disorders involving aberrant protein expression or biological activity.  
 CC The polypeptide and polynucleotide sequences have applications in  
 CC diagnostics, forensics, gene mapping, identification of mutations  
 CC responsible for genetic disorders or other traits to assess biodiversity  
 CC and to produce other types of data and products dependent on DNA and  
 CC amino acid sequences. ABG00010-ABG30377 represent novel human  
 CC diagnostic amino acid sequences.  
 CC Note: The sequence data for this patent did not appear in the printed  
 CC specification, but was obtained in electronic format directly from WIPO  
 CC at ftp.wipo.int/pub/published\_pct\_sequences.  
 CC  
 XX  
 SQ Sequence 297 AA;  
 Query Match 47.8%; Score 43; DB 22; Length 297;  
 Best Local Similarity 66.7%; Pred. No. 45;  
 Matches 8; Conservative 1; Mismatches 3; Indels 0; Gaps 0;  
 QY 4 GRDNDRDPSPK 15  
 ||:|||||  
 Db 58 GRGTHDRDPSEK 69  
 RESULT 18  
 ABB68139  
 ID ABB68139 standard; Protein: 444 AA.  
 AC ABB68139;  
 DT 26-MAR-2002 (first entry)  
 XX  
 DE Drosophila melanogaster polypeptide SEQ ID NO 31209.  
 XX  
 KW Drosophila; developmental biology; cell signalling; insecticide;  
 KW pharmaceutical.  
 XX  
 OS Drosophila melanogaster.  
 XX  
 PN WO200171042-A2.  
 XX  
 PD 27-SEP-2001.  
 XX  
 PF 23-MAR-2001; 2001WO-US09231.  
 XX  
 PR 23-MAR-2000; 2000US-191637P.  
 XX  
 PR 11-JUL-2000; 2000US-0614150.  
 XX  
 PA (PEKE ) PE CORP NY.  
 XX  
 PI Venter JC, Adams M, Li PWD, Myers EW;  
 XX  
 DR WPI; 2001-656860/75.  
 DR N-PSDB; ABL12242.  
 XX  
 PT New isolated nucleic acid detection reagent for detecting 1000 or more  
 PT genes from Drosophila and for elucidating cell signalling and cell-cell  
 PT interactions -  
 XX  
 PS Disclosure; SEQ ID NO 31209; 21bp + Sequence Listing; English.

XX  
 CC The invention relates to an isolated nucleic acid detection reagent  
 CC capable of detecting 1000 or more genes from Drosophila. The invention is  
 CC useful in developmental biology and in elucidating cell signalling and  
 CC cell-cell interactions in higher eukaryotes for the development of  
 CC insecticides, therapeutics and pharmaceutical drugs. The invention  
 CC discloses genomic DNA sequences (AB116176-AB130511), expressed DNA  
 CC sequences (AB101840-AB116175) and the encoded proteins  
 CC (AB57737-AB572072).  
 CC The sequence data for this patent did not form part of the printed  
 CC specification, but was obtained in electronic format directly from WIPO  
 CC at ftp.wipo.int/pub/published\_pct\_sequences.  
 CC  
 XX  
 SQ Sequence 444 AA;  
 Query Match 47.8%; Score 43; DB 22; Length 444;  
 Best Local Similarity 61.5%; Pred. No. 70;  
 Matches 8; Conservative 0; Mismatches 5; Indels 0; Gaps 0;  
 QY 2 GYGRDNDRDPSPK 14  
 |||||  
 Db 402 GYGRDNDRDPSPK 414  
 RESULT 19  
 AAY94944  
 ID AAY94944 standard; Protein: 451 AA.  
 AC AAY94944;  
 DT 16-JUN-2000 (first entry)  
 XX  
 DE Human secreted protein clone b157\_16 protein sequence SEQ ID NO:94.  
 XX  
 KW Human; secreted protein; immunostimulant; immunosuppressant; virucide;  
 KW antibacterial; antifungal; cytostatic; antiinflammatory; dermatological;  
 KW antidiabetic; antidiarrheal; antidiarrheal; antidiarrheal; protozoacide;  
 KW antihypertensive; immune deficiency; severe combined immunodeficiency; SCID;  
 KW infection; HIV; hepatitis; malaria; autoimmune disorder; systemic lupus;  
 KW connective tissue disease; multiple sclerosis; erythematosis;  
 KW rheumatoid arthritis; autoimmune pulmonary inflammation; asthma;  
 KW Guillain-Barre syndrome; autoimmune thyroiditis; myasthenia gravis;  
 KW insulin dependent diabetes mellitus; graft-versus-host-disease;  
 KW autoimmune inflammatory eye disease; allergy.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO200009552-A1.  
 XX  
 PD 24-FEB-2000.  
 XX  
 PF 13-AUG-1999; 99WO-US18298.  
 XX  
 PR 14-AUG-1998; 98US-0096622.  
 PR 17-AUG-1998; 98US-0096815.  
 PR 04-SEP-1998; 98US-0099229.  
 PR 23-OCT-1998; 98US-0105368.  
 PR 08-JAN-1999; 99US-0115234.  
 PR 12-FEB-1999; 99US-0119931.  
 PR 18-FEB-1999; 99US-0120575.  
 PR 30-APR-1999; 99US-0132020.  
 PR 11-AUG-1999; 99US-0096622.  
 XX  
 PA (GBMY ) GENETICS INST INC.  
 XX  
 PI Jacobs K, McCoy JM, LaValle ER, Collins-Racie LA, Evans C;  
 PI Merberg D, Treacy M, Agostino MJ, Steininger RJ, Spaulding V;  
 PI Wong GG, Clark HF, Fechtel K;  
 XX  
 DR WPI; 2000-205979/18.  
 XX  
 PT New polynucleotides encoding secreted proteins, which may have e.g.  
 PT nutritional, chemokine, immune stimulating or suppressing,

PT hematopoiesis regulating, tissue growth, activin/inhibin  
PT antiinflammatory or tumor inhibition activity -  
XX  
XX  
PS Claim 103; Page 562-563; 641pp; English.  
CC AAA16618 to AAA16637 encode the human secreted proteins given in  
CC AAY94898 to AAY94980, isolated from human adult brain, adult thyroid,  
CC adult retina, foetal carcinoma, adult blood, adult neural, foetal kidney,  
CC adult placenta, adult testis, whole embryo, adult cartilage, kidney,  
CC foetal brain, adult thymus, foetal placenta, adult uterus, adult tumour,  
CC and adult bladder, cDNA libraries. The polynucleotides and proteins are  
CC predicted to have biological activities which would make them suitable  
CC for treating, preventing or ameliorating medical conditions in humans  
CC and animals. The polynucleotides can be used as markers for tissues in  
CC which the protein is preferentially expressed, as molecular weight  
CC markers on Southern gels, and as chromosome markers or tags to identify  
CC chromosomes or to map gene positions. The proteins can be used in the  
CC treatment of immune deficiencies and disorders, such as severe combined  
CC immunodeficiency (SCID), as well as viral, bacterial, fungal and other  
CC infections. These infections include human immunodeficiency virus (HIV),  
CC hepatitis, herpesviruses, mycobacteria, leishmania spp., malaria and  
CC candidiasis. The proteins can be used to treat autoimmune disorders such  
CC as connective tissue disease, multiple sclerosis, systemic lupus  
CC erythematosus, rheumatoid arthritis, autoimmune pulmonary inflammation,  
CC Guillain-Barre syndrome, autoimmune thyroiditis, insulin dependent  
CC diabetes mellitus, myasthenia gravis, graft-versus-host disease and  
CC autoimmune inflammatory eye disease. The proteins can also be used to  
CC treat allergic conditions, such as asthma. AAA16598 to AAA16774 represent  
CC probes for the human secreted proteins from the present invention.  
XX  
SQ Sequence 451 AA;  
Query Match 47.8%; Score 43; DB 21; Length 451;  
Best Local Similarity 66.7%; Pred. No. 72;  
Matches 8; Conservative 1; Mismatches 3; Indels 0; Gaps 0;  
QY 4 GRDNDNRDPSRK 15  
||:|||||  
Db 165 GRGTHDRDPSRK 176  
RESULT 20  
AAM38966  
ID AAM38966 standard; Protein; 477 AA.  
XX  
AC AAM38966;  
XX  
DT 22-OCT-2001 (first entry)  
XX  
DE Human polypeptide SEQ ID NO 2111.  
XX  
KW Human; noctropic; immunosuppressant; cyrostatic; gene therapy; cancer;  
KW peripheral nervous system; neuropathy; central nervous system; CNS;  
KW Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;  
KW amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;  
KW chemokine; thrombolytic; drug screening; arthritis; inflammation;  
KW leukaemia.  
XX  
XX Homo sapiens.  
XX  
PN WO200153312-A1.  
XX  
PD 26-JUL-2001.  
XX  
PF 26-DEC-2000; 2000WO-US34263.  
XX  
XX 21-JAN-2000; 2000US-0488725.  
XX 25-APR-2000; 2000US-0552317.  
XX 09-JUL-2000; 2000US-0598042.  
XX 19-JUL-2000; 2000US-0620312.  
XX 03-AUG-2000; 2000US-0653450.  
XX 14-SEP-2000; 2000US-0662191.  
XX 19-OCT-2000; 2000US-0693036.

PR 29-NOV-2000; 2000US-0727344.  
XX  
XX (HYSE-) HYSEQ INC.  
XX  
PI Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;  
PI Wang U, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J;  
PI Zhao QA, Zhou F, Goodrich R, Drmanac RT;  
XX WPI: 2001-442253/47.  
DR N-PSDB; AAI58122.  
XX  
PT Novel nucleic acids and polypeptides, useful for treating disorders  
PT such as central nervous system injuries -  
XX  
PS Example 3; SEQ ID NO 2111; 10078pp; English.  
XX  
CC The invention relates to human nucleic acids (AAI57798-AAI61369) and  
CC the encoded polypeptides (AAM38642-AAM42213) with noctropic,  
CC immunosuppressant and cyostatic activity. The polynucleotides are useful  
CC in gene therapy. A composition containing a polypeptide or polynucleotide  
CC of the invention may be used to treat diseases of the peripheral nervous  
CC system, such as peripheral nervous injuries, peripheral neuropathy and  
CC localised neuropathies and central nervous system diseases, such as  
CC Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic  
CC lateral sclerosis, and Shy-Drager Syndrome. Other uses include the  
CC utilisation of the activities such as: Immune system suppression,  
CC Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic  
CC and thrombolytic activity, cancer diagnosis and therapy, drug screening,  
CC assays for receptor activity, arthritis and inflammation, leukemias and  
CC C.N.S disorders.  
CC Note: The sequence data for this patent did not form part of the printed  
CC specification.  
XX  
SQ Sequence 477 AA;  
Query Match 47.8%; Score 43; DB 22; Length 477;  
Best Local Similarity 66.7%; Pred. No. 76;  
Matches 8; Conservative 1; Mismatches 3; Indels 0; Gaps 0;  
QY 4 GRDNDNRDPSRK 15  
||:|||||  
Db 165 GRGTHDRDPSRK 176  
RESULT 21  
ABG10605  
ID ABG10605 standard; Protein; 515 AA.  
XX  
AC ABG10605;  
XX  
DT 13-FEB-2002 (first entry)  
XX  
DE Novel human diagnostic protein #10596.  
XX  
KW Human; chromosome mapping; gene mapping; gene therapy; forensic;  
KW food supplement; medical imaging; diagnostic; genetic disorder.  
XX  
XX Homo sapiens.  
XX  
PN WO200175067-A2.  
XX  
PD 11-OCT-2001.  
XX  
PF 30-MAR-2001; 2001WO-US08631.  
XX  
XX 31-MAR-2000; 2000US-0540217.  
XX 23-AUG-2000; 2000US-0649167.  
XX  
PA (HYSE-) HYSEQ INC.  
XX  
PI Drmanac RT, Liu C, Tang YT;  
XX WPI: 2001-639362/73.

DR N-PSDB; AAS74792.  
XX New isolated polynucleotide and encoded polypeptides, useful in  
PT diagnostics, forensics, gene mapping, identification of mutations  
PT responsible for genetic disorders or other traits and to assess  
PT biodiversity -  
XX Claim 20; SEQ ID No 40964; 103bp; English.  
PS  
XX  
CC The invention relates to isolated polynucleotide (I) and  
CC polypeptide (II) sequences. (I) is useful as hybridization probes,  
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome  
CC and gene mapping, and in recombinant production of (II). The  
CC polynucleotides are also used in diagnostics as expressed sequence tags  
CC for identifying expressed genes. (I) is useful in gene therapy techniques  
CC to restore normal activity of (II) or to treat disease states involving  
CC (II). (II) is useful for generating antibodies against it, detecting or  
CC quantitating a polypeptide in tissue, as molecular weight markers and as  
CC a food supplement. (II) and its binding partners are useful for treating  
CC imaging of sites expressing (II). (I) and (II) are useful for treating  
CC disorders involving aberrant protein expression or biological activity.  
CC The polypeptide and polynucleotide sequences have applications in  
CC diagnostics, forensics, gene mapping, identification of mutations  
CC responsible for genetic disorders or other traits to assess biodiversity  
CC and to produce other types of data and products dependent on DNA and  
CC amino acid sequences. ABG00010-ABG30377 represent novel human  
CC diagnostic amino acid sequences of the invention.  
CC Note: The sequence data for this patent did not appear in the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pct\_sequences.  
XX  
SQ Sequence 515 AA;  
XX  
Query Match 47.8%; Score 43; DB 22; Length 515;  
Best Local Similarity 66.7%; Pred. No. 83;  
Matches 8; Conservative 1; Mismatches 3; Indels 0; Gaps 0;  
QY 4 GRDNDRDPSRK 15  
||:|||||  
Db 165 GRGTHDRDPSEK 176  
RESULT 22  
ABG10607  
ID ABG10607 standard; Protein; 809 AA.  
XX  
AC ABG10607;  
XX  
DT 13-FEB-2002 (first entry)  
XX  
DE Novel human diagnostic protein #10598.  
XX  
KW Human; chromosome mapping; gene mapping; gene therapy; forensic;  
KW food supplement; medical imaging; diagnostic; genetic disorder.  
XX  
OS Homo sapiens.  
XX  
PN WO200175067-A2.  
XX  
PD 11-OCT-2001.  
XX  
PF 30-MAR-2001; 2001WO-US08631.  
XX  
PR 31-MAR-2000; 2000US-0540217.  
PR 23-AUG-2000; 2000US-0649167.  
XX  
PA (HYSE-) HYSEQ INC.  
XX  
PI Drmanac RT, Liu C, Tang YT;  
XX  
DR WPI, 2001-639362/73.  
DR N-PSDB; AAS74794.  
XX

PT New isolated polynucleotide and encoded polypeptides, useful in  
PT diagnostics, forensics, gene mapping, identification of mutations  
PT responsible for genetic disorders or other traits and to assess  
PT biodiversity -  
XX Claim 20; SEQ ID No 40966; 103bp; English.  
PS  
XX  
CC The invention relates to isolated polynucleotide (I) and  
CC polypeptide (II) sequences. (I) is useful as hybridization probes,  
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome  
CC and gene mapping, and in recombinant production of (II). The  
CC polynucleotides are also used in diagnostics as expressed sequence tags  
CC for identifying expressed genes. (I) is useful in gene therapy techniques  
CC to restore normal activity of (II) or to treat disease states involving  
CC (II). (II) is useful for generating antibodies against it, detecting or  
CC quantitating a polypeptide in tissue, as molecular weight markers and as  
CC a food supplement. (II) and its binding partners are useful in medical  
CC imaging of sites expressing (II). (I) and (II) are useful for treating  
CC disorders involving aberrant protein expression or biological activity.  
CC The polypeptide and polynucleotide sequences have applications in  
CC diagnostics, forensics, gene mapping, identification of mutations  
CC responsible for genetic disorders or other traits to assess biodiversity  
CC and to produce other types of data and products dependent on DNA and  
CC amino acid sequences. ABG00010-ABG30377 represent novel human  
CC diagnostic amino acid sequences of the invention.  
CC Note: The sequence data for this patent did not appear in the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pct\_sequences.  
XX  
SQ Sequence 809 AA;  
XX  
Query Match 47.8%; Score 43; DB 22; Length 809;  
Best Local Similarity 66.7%; Pred. No. 1.4e+02;  
Matches 8; Conservative 1; Mismatches 3; Indels 0; Gaps 0;  
QY 4 GRDNDRDPSRK 15  
||:|||||  
Db 165 GRGTHDRDPSEK 176  
RESULT 23  
ABG29958  
ID ABG29958 standard; Protein; 205 AA.  
XX  
AC ABG29958;  
XX  
DT 18-FEB-2002 (first entry)  
XX  
DE Novel human diagnostic protein #29949.  
XX  
KW Human; chromosome mapping; gene mapping; gene therapy; forensic;  
KW food supplement; medical imaging; diagnostic; genetic disorder.  
XX  
OS Homo sapiens.  
XX  
PN WO200175067-A2.  
XX  
PD 11-OCT-2001.  
XX  
PF 30-MAR-2001; 2001WO-US08631.  
XX  
PR 31-MAR-2000; 2000US-0540217.  
PR 23-AUG-2000; 2000US-0649167.  
XX  
PA (HYSE-) HYSEQ INC.  
XX  
PI Drmanac RT, Liu C, Tang YT;  
XX  
DR WPI, 2001-639362/73.  
DR N-PSDB; AAS94145.  
XX  
PT New isolated polynucleotide and encoded polypeptides, useful in  
PT diagnostics, forensics, gene mapping, identification of mutations

PT responsible for genetic disorders or other traits and to assess  
PT biodiversity -  
XX  
PS Claim 20; SEQ ID No 60317; 103pp; English.  
XX  
CC The invention relates to isolated polynucleotide (I) and  
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,  
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome  
CC and gene mapping, and in recombinant production of (II). The  
CC polynucleotides are also used in diagnostics as expressed sequence tags  
CC for identifying expressed genes. (I) is useful in gene therapy techniques  
CC to restore normal activity of (II) or to treat disease states involving  
CC (II). (II) is useful for generating antibodies against it, detecting or  
CC quantitating a polypeptide in tissue, as molecular weight markers and as  
CC a food supplement. (II) and its binding partners are useful in medical  
CC imaging of sites expressing (II). (I) and (II) are useful for treating  
CC disorders involving aberrant protein expression or biological activity.  
CC The polypeptide and polynucleotide sequences have applications in  
CC diagnostics, forensics, gene mapping, identification of mutations  
CC responsible for genetic disorders or other traits to assess biodiversity  
CC and to produce other types of data and products dependent on DNA and  
CC amino acid sequences. ABG00010-ABG30377 represent novel human  
CC diagnostic amino acid sequences of the invention.  
CC Note: The sequence data for this patent did not appear in the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pct\_sequences.  
XX  
SQ Sequence 205 AA;  
  
Query Match 46.7%; Score 42; DB 22; Length 205;  
Best Local Similarity 70.0%; Pred. No. 43;  
Matches 7; Conservative 2; Mismatches 1; Indels 0; Gaps 0;  
  
QY 4 GRDNDNDRPPS 13  
| | | : | | | | |  
Db 28 GRDPSDRDPA 37  
  
RESULT 24  
ID ABB61193 standard; Protein: 215 AA.  
XX  
AC ABB61193;  
XX  
DT 26-MAR-2002 (first entry)  
XX  
DE Drosophila melanogaster polypeptide SEQ ID NO 10371.  
XX  
KW Drosophila; developmental biology; cell signalling; insecticide;  
KW pharmaceutical.  
XX  
OS Drosophila melanogaster.  
XX  
FN WO200171042-A2.  
XX  
PD 27-SEP-2001.  
XX  
PF 23-MAR-2001; 2001WO-US09231.  
XX  
PR 23-MAR-2000; 2000US-191637P.  
PR 11-JUL-2000; 2000US-0614150.  
XX  
PA (PEKE ) PE CORP NY.  
XX  
PI Venter JC, Adams M, Li PWD, Myers EW;  
XX  
DR WPI; 2001-656860/75.  
DR N-PSDB; ABL05296.  
XX  
PT New isolated nucleic acid detection reagent for detecting 1000 or more  
PT genes from Drosophila and for elucidating cell signalling and cell-cell  
PT interactions -  
XX

PS Disclosure; SEQ ID NO 10371; 21pp + Sequence Listing; English.  
XX  
CC The invention relates to an isolated nucleic acid detection reagent  
CC capable of detecting 1000 or more genes from Drosophila. The invention is  
CC useful in developmental biology and in elucidating cell signalling and  
CC cell-cell interactions in higher eukaryotes for the development of  
CC insecticides, therapeutics and pharmaceutical drugs. The invention  
CC discloses genomic DNA sequences (ABL01840-ABL16175), expressed DNA  
CC sequences (ABL01840-ABL16175) and the encoded proteins  
CC (ABR57737-ABR72072).  
CC The sequence data for this patent did not form part of the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pct\_sequences.  
XX  
SQ Sequence 215 AA;  
  
Query Match 46.7%; Score 42; DB 22; Length 215;  
Best Local Similarity 50.0%; Pred. No. 46;  
Matches 7; Conservative 3; Mismatches 4; Indels 0; Gaps 0;  
  
QY 1 VGYGRDNDNDRPPSR 14  
: | | : | | | | |  
Db 14 IPRGEBDEMRDPSR 27  
  
RESULT 25  
ID ABG29954 standard; Protein: 254 AA.  
XX  
AC ABG29954;  
XX  
DT 18-FEB-2002 (first entry)  
XX  
DE Novel human diagnostic protein #29945.  
XX  
KW Human; chromosome mapping; gene mapping; gene therapy; forensic;  
KW food supplement; medical imaging; diagnostic; genetic disorder.  
XX  
OS Homo sapiens.  
XX  
FN WO200175067-A2.  
XX  
PD 11-OCT-2001.  
XX  
PF 30-MAR-2001; 2001WO-US08631.  
XX  
PR 31-MAR-2000; 2000US-0540217.  
PR 23-AUG-2000; 2000US-0649167.  
XX  
PA (HYSE-) HYSEQ INC.  
XX  
PI Drmanac RT, Liu C, Tang YT;  
XX  
DR WPI; 2001-639362/73.  
DR N-PSDB; AAS94141.  
XX  
PT New isolated polynucleotide and encoded polypeptides, useful in  
PT diagnostics, forensics, gene mapping, identification of mutations  
PT responsible for genetic disorders or other traits and to assess  
PT biodiversity -  
XX  
PS Claim 20; SEQ ID No 60313; 103pp; English.  
XX  
CC The invention relates to isolated polynucleotide (I) and  
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,  
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome  
CC and gene mapping, and in recombinant production of (II). The  
CC polynucleotides are also used in diagnostics as expressed sequence tags  
CC for identifying expressed genes. (I) is useful in gene therapy techniques  
CC to restore normal activity of (II) or to treat disease states involving  
CC (II). (II) is useful for generating antibodies against it, detecting or  
CC quantitating a polypeptide in tissue, as molecular weight markers and as  
CC a food supplement. (II) and its binding partners are useful in medical

CC imaging of sites expressing (II). (I and (II) are useful for treating  
CC disorders involving aberrant protein expression or biological activity.  
CC The polypeptide and polynucleotide sequences have applications in  
CC diagnostics, forensics, gene mapping, identification of mutations  
CC responsible for genetic disorders or other traits to assess biodiversity  
CC and to produce other types of data and products dependent on DNA and  
CC amino acid sequences. ABG00010-ABG30377 represent novel human  
CC diagnostic amino acid sequences of the invention.  
CC Note: The sequence data for this patent did not appear in the printed  
CC specification, but was obtained in electronic format directly from WHO  
CC at [ftp.who.int/pub/published\\_pct\\_sequences](http://who.int/pub/published_pct_sequences).

CC The polypeptide and polynucleotide sequences have applications in  
CC diagnostics, forensics, gene mapping, identification of mutations  
CC responsible for genetic disorders or other traits to assess biodiversity  
CC and to produce other types of data and products dependent on DNA and  
CC amino acid sequences. ABG00010-ABG30377 represent novel human  
CC diagnostic amino acid sequences of the invention.  
CC Note: The sequence data for this patent did not appear in the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at [ftp.wipo.int/pub/published\\_pct\\_sequences](http://wipo.int/pub/published_pct_sequences).  
XX  
50 Sequence 361 AA;

[illegible]

PR	08-NOV-2000;	2000US-0246527.
PR	08-NOV-2000;	2000US-0246528.
PR	08-NOV-2000;	2000US-0246532.
PR	08-NOV-2000;	2000US-0246532.
PR	08-NOV-2000;	2000US-0246609.
PR	08-NOV-2000;	2000US-0246610.
PR	08-NOV-2000;	2000US-0246611.
PR	08-NOV-2000;	2000US-0246613.
PR	17-NOV-2000;	2000US-0249207.
PR	17-NOV-2000;	2000US-0249208.
PR	17-NOV-2000;	2000US-0249209.
PR	17-NOV-2000;	2000US-0249210.
PR	17-NOV-2000;	2000US-0249211.
PR	17-NOV-2000;	2000US-0249212.
PR	17-NOV-2000;	2000US-0249213.
PR	17-NOV-2000;	2000US-0249214.
PR	17-NOV-2000;	2000US-0249215.
PR	17-NOV-2000;	2000US-0249216.
PR	17-NOV-2000;	2000US-0249217.
PR	17-NOV-2000;	2000US-0249218.
PR	17-NOV-2000;	2000US-0249219.
PR	17-NOV-2000;	2000US-0249220.
PR	17-NOV-2000;	2000US-0249221.
PR	17-NOV-2000;	2000US-0249222.
PR	17-NOV-2000;	2000US-0249223.
PR	17-NOV-2000;	2000US-0249224.
PR	17-NOV-2000;	2000US-0249225.
PR	17-NOV-2000;	2000US-0249226.
PR	17-NOV-2000;	2000US-0249227.
PR	17-NOV-2000;	2000US-0249228.
PR	17-NOV-2000;	2000US-0249229.
PR	17-NOV-2000;	2000US-0249230.
PR	01-DEC-2000;	2000US-0250160.
PR	01-DEC-2000;	2000US-0250391.
PR	05-DEC-2000;	2000US-0251030.
PR	05-DEC-2000;	2000US-0251988.
PR	05-DEC-2000;	2000US-0256719.
PR	06-DEC-2000;	2000US-0251749.
PR	08-DEC-2000;	2000US-0251856.
PR	08-DEC-2000;	2000US-0251856.
PR	08-DEC-2000;	2000US-0251869.
PR	08-DEC-2000;	2000US-0251989.
PR	08-DEC-2000;	2000US-0251990.
PR	11-DEC-2000;	2000US-0254097.
PR	05-JAN-2001;	2001US-0259678.
XX	(HUMA-) HUMAN GENOME SCI INC.	
PA		
XX		
P1	Rosen CA, Barash SC, Ruben SM;	
XX		
DR	WPI; 2001-488783/53.	
DR	N-PSDB; AAS26396.	
XX		
PT	New nucleic acid molecules encoding 461 human secreted proteins for	
PT	diagnosing, preventing, treating or ameliorating medical conditions and	
PT	used as food additives or preservatives -	
PT		
PS	Claim 11; SEQ ID No 1362; 980bp; English.	
XX		
XX		
CC	The invention relates to isolated nucleic acid molecules and their	
CC	encoded secreted proteins. The nucleic acids and proteins are used to	
CC	prevent, treat or ameliorate a medical condition in e.g. humans, mice,	
CC	rabbits, goats, horses, cats, dogs, chickens or sheep. They	
CC	are also used in diagnosing a pathological condition or susceptibility	
CC	to a pathological condition. Antibodies to the proteins can also	
CC	be used in alleviating symptoms associated with the disorders and in	
CC	diagnostic immunoassays e.g. radioimmunoassays or enzyme linked	
CC	immunosorbent assays (ELISA). Disorders which are diagnosed or treated	
CC	include autoimmune diseases e.g. rheumatoid arthritis,	
CC	hyperproliferative disorders e.g. neoplasms of the breast or liver,	
CC	cardiovascular disorders e.g. cardiac arrest, cerebrovascular disorders	
CC	e.g. cerebral ischaemia, anglogenesis, nervous system disorders e.g.	
CC	Alzheimer's disease, infections caused by bacteria, viruses and fungi	
CC	and ocular disorders e.g. corneal infection, and many other	
CC	disorders listed in the specification. The polypeptides can also	
CC	be used to aid wound healing and epithelial cell proliferation, to	
CC	prevent skin aging due to sunburn, to maintain organs before	
CC	transplantation, for supporting cell culture of primary tissues, to	
CC	regenerate tissues and in chemotaxis. The polypeptides can also be used	

CC as a food additive or preservative to increase or decrease storage  
 CC capabilities, fat content, lipid, protein, carboxylate, vitamins,  
 CC minerals, cofactors and other nutritional components. The present  
 CC sequence represents a novel secreted protein of the invention.

Query Match 46.7%; Score 42; DB 22; Length 500;  
 Best Local Similarity 66.7%; Pred. No. 1.2e+02;  
 Matches 8; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

Qy 4 GRDNDPDRPSRK 15  
 Db 188 GRGTHDRPSKK 199

RESULT 28  
 AAB43890  
 ID AAB43890 standard; Protein; 1005 AA.

AC AAB43890;  
 DT 08-FEB-2001 (first entry)

DE Human cancer associated protein sequence SEQ ID NO:1335.

XX Human; cancer associated gene; cancer antigen; detection; cancer;  
 KM diagnosis; cytostatic; proliferative; vulnerability; immunomodulator;  
 KM antidiabetic; antiaesthetic; antirheumatic; antiarthritic; antiviral;  
 KM antiinflammatory; antihypoid; antiallergic; antibacterial; cardiac;  
 KM dermatologic; neuroprotective; thrombolytic; coagulant; neotropic;  
 KM vasotropic; antiprotic; antiangiogenic; gene therapy; inflammation;  
 KM immune disorder; haematopoietic cell disorder; autoimmune disorder;  
 KM allergic reaction; graft versus host disease; organ rejection;  
 KM haemostatic; thrombolytic; cardiovascular disorder; infection;  
 KM neurological disease; drug screening.

XX Homo sapiens.

OS WO20005350-A1.

PN 21-SEP-2000.

PD 08-MAR-2000; 2000WO-US05882.

PF 12-MAR-1999; 99US-0124270.

PR (HUMA-) HUMAN GENOME SCI INC.

PA Rosen CA, Ruben SM;

XX WPI; 2000-58753/55.

DR N-PSDB; AAC78099.

PT Novel isolated nucleic acids comprising sequences encoding peptides  
 useful for treating or diagnosing e.g. cancer -

PS Claim 11; Page 1987-1991; 2352pp; English.

XX AAC77607 to AAC78448 encode the human cancer associated proteins given  
 CC in AAB43398 to AAB44233. The proteins can have activities based on the  
 CC tissues and cells the genes are expressed in. Example of activities  
 CC include: cytostatic; proliferative; vulnerary; immunomodulator;  
 CC antidiabetic; antiaesthetic; antirheumatic; antiarthritic;  
 CC antiinflammatory; antihypoid; antiallergic; antibacterial; antiviral;  
 CC dermatologic; neuroprotective; cardiac; thrombolytic; coagulant;  
 CC neotropic; vasotropic; antiprotic and antiangiogenic. The  
 CC polynucleotides and polypeptides can be used for preventing, treating or  
 CC ameliorating medical conditions and diagnosing pathological conditions.  
 CC Polynucleotides, polypeptides, antibodies, agonists and antagonists from  
 CC the present invention may be used to treat immune disorders by activating  
 CC or inhibiting the proliferation, differentiation or mobilisation of  
 CC immune cells, to treat disorders of haematopoietic cells, autoimmune  
 CC disorders, allergic reactions, graft versus host disease and organ  
 CC rejection, modulate haemostatic or thrombolytic activity, modulate

CC inflammation, cancers, cardiovascular disorders, neurological disease and  
 CC bacterial or viral infections. The peptides, nucleotides, antibodies,  
 CC agonists and antagonists may be also be used in drug screens. AAC78449 to  
 CC AAC78457 and AAB44240 represent sequences used in the exemplification of  
 CC the present invention.

XX Sequence 1005 AA;

Query Match 46.7%; Score 42; DB 21; Length 1005;  
 Best Local Similarity 54.5%; Pred. No. 2.6e+02;  
 Matches 6; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

Qy 5 RDDNDPDRPSRK 15  
 Db 839 REENDKDPERE 849

RESULT 29  
 ABB93183  
 ID ABB93183 standard; Protein; 1179 AA.

AC ABB93183;

DT 31-MAY-2002 (first entry)

DE Herbicidally active polypeptide SEQ ID NO 2394.

XX Herbicidal; plant; agriculture; herbicide.

XX Arabidopsis thaliana.

OS WO200210210-A2.

PN 07-FEB-2002.

PD 28-AUG-2001; 2001WO-EP09892.

PF 28-AUG-2001; 2001WO-EP09892.

PR (FARB ) BAYER AG.

PA Tietjen K, Weidner M;

XX WPI; 2002-269010/31.

PT Identifying plant target proteins for herbicidally active compounds,  
 comprising aligning and comparing nucleic acid or amino acid sequences  
 from plant with nucleic acid or amino acid sequences from non-plant  
 organisms -

PS Claim 5; SEQ ID NO 2394; 261pp + Sequence Listing; English.

XX The invention relates to identifying target proteins  
 CC (ABB90790-ABB94016) for herbicidally active compounds, comprising  
 CC aligning and comparing nucleic acid or amino acid sequences from plant  
 CC with nucleic acid or amino acid sequences from non-plant organisms using  
 CC suitable search parameters, where plant sequences having an E-value  
 CC greater by a factor of 3 than the E-value of most similar non-plant  
 CC sequences are selected. The polypeptides or nucleic acids encoding them  
 CC are useful for identifying modulators. The identified modulators are  
 CC useful as herbicides.

XX Sequence 1179 AA;

Query Match 46.7%; Score 42; DB 23; Length 1179;  
 Best Local Similarity 42.9%; Pred. No. 3.1e+02;  
 Matches 6; Conservative 4; Mismatches 4; Indels 0; Gaps 0;

Qy 3 YGRDNDPDRPSRK 16  
 Db 1019 YGEDEDDPPKNN 1032

```

RESULT 30
AAW31867
ID AAW31867 standard; Protein; 1382 AA.
XX
XX
AC AAW31867;
XX
DT 27-APR-1998 (first entry)
XX
DE Human metastasis-associated antigen p150.
XX
KW Antigen; p150; metastasis; tumour; cancer; diagnosis; therapy;
KM human.
XX
XX Homo sapiens.
OS
PN W09741221-A1.
XX
XX 06-NOV-1997.
PD
XX
XX 18-APR-1997; 97WO-EP01963.
PF
XX
PR 28-FEB-1997; 97GB-0004161.
PR 29-APR-1996; 96US-0016487.
XX
PA (NOVS ) NOVARTIS AG.
XX
XX Bachmann F;
PI
XX WPI; 1997-549725/50.
DR
DR N-PSDB; AAT93627.
XX
XX Mouse and human metastasis-associated p150-derived antigens - useful
PT for raising antibodies for therapy and diagnosis of tumours
XX
XX Claim 3; Page 86-91; 11pp; English.
PS
XX
XX This protein comprises a novel human 150 kDa cell surface
CC metastasis-associated antigen, designated p150. It is the
CC homologue of a novel murine p150 metastasis-associated antigen (see
CC AAW31866). Its amino acid sequence was deduced from an isolated cDNA
CC clone (see AAT93627). A claimed method for determining the
CC metastatic potential of a tumour cell comprises assessing the level
CC of p150 expression in the cell, a higher level of p150 expression
CC being indicative of a lower state of differentiation and
CC consequently a higher metastatic potential. Immunogenic
CC determinants of p150 and antibodies raised against them are
CC useful in the diagnosis and treatment of metastatic potential in
CC tumours.
XX
XX Sequence 1382 AA;
SQ
Query Match 46.7%; Score 42; DB 18; Length 1382;
Best Local Similarity 54.5%; Pred. No. 3.7e+02;
Matches 6; Conservative 4; Mismatches 1; Indels 0; Gaps 0;
QY 5 RDDNDRDPSPK 15
|:|:|:|:|:|
Db 1216 REENDXDPERE 1226

```

```

CS Mus musculus.
XX
XX W0200188188-A2.
XX
XX 22-NOV-2001.
PD
XX
XX 18-MAY-2001; 2001MO-JP04192.
PF
XX
XX 18-MAY-2000; 2000JP-0145977.
PR
XX
XX (UYN1-) UNIV NIHON SCHOOL JURIDICAL PERSON.
PA
XX
XX Ishikawa K, Asai S, Takahashi Y, Nagata T, Ishii Y;
PI
XX
XX WPI; 2002-034733/04.
DR
DR N-PSDB; ABI99479.
XX
XX Examining the ischemic condition (e.g. occlusive ischemia) by measuring
PT expression levels of particular genes defined in the specification or
PT by determining the expression profile of a gene group comprising these
PT genes -
XX
XX Claim 2; Page 1230-1241; 2690pp; English.
PS
XX
XX The present invention describes a method for examining ischemic
CC conditions, comprising measuring the expression levels of particular
CC genes (I) in a test sample or determining the expression profile of a
CC gene group in the sample comprising genes selected from (I). The method
CC is useful for examining the ischemic condition (e.g. compressive
CC ischemia, occlusive ischemia or vasospastic ischemia) by measuring
CC expression levels of particular genes (ABI99202 to ABI99912, encoding
CC the protein sequences in ABB57020 to ABB57374) or by determining the
CC expression profile of a gene group comprising these genes. The
CC expression levels or expression profiles produced by these genes are
CC used as an indicator when screening for ischemic condition-improving
CC drugs or therapeutics for ischemic diseases. ABI99913 and ABI99914
CC represent PCR primers for a mouse ischemic condition related sequence,
CC which are used in the exemplification of the present invention.
XX
XX Sequence 2559 AA;
SQ
Query Match 46.7%; Score 42; DB 23; Length 2559;
Best Local Similarity 70.0%; Pred. No. 7.5e+02;
Matches 7; Conservative 1; Mismatches 2; Indels 0; Gaps 0;
QY 3 YGRDNDRDP 12
|:|:|:|:|
Db 630 YGRDNDYDP 639

```

```

RESULT 31
ABB57181
ID ABB57181 standard; Protein; 2559 AA.
XX
XX ABB57181;
AC
XX
DT 07-MAR-2002 (first entry)
XX
DE Mouse ischemic condition related protein sequence SEQ ID NO:440.
XX
XX Mouse; ischemia; compressive ischemia; occlusive ischemia;
KM vasospastic ischemia; ischemic condition; ischemic disease.
XX
XX

```

```

RESULT 32
AA59894
ID AA59894 standard; Protein; 127 AA.
XX
XX AA59894;
AC
XX
DT 19-JAN-2000 (first entry)
XX
DE Human normal pancreas tissue derived protein 2.
XX
XX Human; pancreas; cancer; treatment; anticancer; cytostatic;
KW gene therapy; ESR; expressed sequence tag.
XX
XX Homo sapiens.
OS
XX
XX DE19818598-A1.
PN
XX
XX 21-OCT-1999.
PD
XX
XX 19-APR-1998; 98DE-1018598.
PF
XX
XX 19-APR-1998; 98DE-1018598.
PR
XX
XX

```



PA (META-) METAGEN GES GENOMFORSCHUNG MBH.  
 XX  
 PI Rosenthal A, Specht T, Hinzmann B, Schmitt A, Pilarsky C, Dahl E;  
 XX WPI: 1999-592039/51.  
 DR N-PSDB: AA241409.  
 XX  
 PT New nucleic acid sequences expressed in normal pancreatic tissues, and  
 PT derived polypeptides, for treatment of pancreatic cancer and  
 PT identification of therapeutic agents  
 XX  
 PS Claim 23; Page 71; 92pp; German.  
 XX  
 CC This invention describes novel cDNA sequences (A) that are highly  
 CC expressed in normal human pancreatic tissue and which have anticancer  
 CC and cytostatic activity. (A) are used (i) for recombinant expression of  
 CC polypeptides (B) and (ii) to isolate complete genes. (B) are used (i) to  
 CC identify agents suitable for treatment of pancreatic cancer; (ii)  
 CC directly for treating this form of cancer (including expression from  
 CC gene therapy vectors) and (iii) for generation of specific antibodies.  
 CC (A) are identified by assembling ESTs (expressed sequence tags) from a  
 CC particular tissue type before comparison of expression patterns. This  
 CC allows a significantly longer fragment of the gene to be revealed, so  
 CC should reduce the number of failures associated with the fact that ESTs  
 CC from different libraries may represent different parts of the same  
 CC unknown gene, distorting the estimated frequency of occurrence in a  
 CC particular tissue. AAY59893-Y59920 represent protein fragments encoded  
 CC from the human normal pancreatic tissue derived cDNA fragments  
 CC represented in AA241386-Z41423.  
 CC  
 XX  
 SQ Sequence 127 AA;  
 Query Match 45.6%; Score 41; DB 20; Length 127;  
 Best Local Similarity 50.0%; Pred. No. 37;  
 Matches 7; Conservative 4; Mismatches 3; Indels 0; Gaps 0;  
 QY 3 YGRDNDRDRPSKRN 16  
 : ||| : ||| :  
 Db 87 HGVDDEDPAPKRD 100  
 RESULT 33  
 AAU87108  
 ID AAU87108 standard; Protein: 127 AA.  
 XX  
 AC AAU87108;  
 XX  
 DT 05-JUN-2002 (first entry)  
 XX  
 DE Novel central nervous system protein #18.  
 XX  
 KM Central nervous system; CNS; autoimmune disease; rheumatoid arthritis;  
 KM hyperproliferative disorder; neoplasm; cardiovascular disorder;  
 KM cardiac arrest; cerebrovascular disorder; ischemia; angiogenesis;  
 KM nervous system disorder; Alzheimer's disease; AIDS; ocular disorder;  
 KM acquired immunodeficiency virus; dysphagia; gastrointestinal disorder;  
 KM adenocarcinoma; reproductive system disorder; testicular feminisation;  
 KM endocrine disorder; diabetes; cancer; leukaemia; neovascularisation;  
 KM respiratory disorder; renal disorder; kidney failure; blood disorder;  
 KM myocardial infarction; wound healing; cell proliferation; skin aging;  
 KM food additive; food preservative; gene therapy.  
 XX  
 OS Homo sapiens.  
 XX  
 PN W0200155318-A2.  
 XX  
 PD 02-AUG-2001.  
 XX  
 PF 17-JAN-2001; 2001WO-US01332.  
 XX  
 PR 31-JAN-2000; 2000US-0179065.  
 PR 04-FEB-2000; 2000US-0180628.  
 PR 24-FEB-2000; 2000US-0184664.

PR 02-MAR-2000; 2000US-0186350.  
 PR 16-MAR-2000; 2000US-0189874.  
 PR 17-MAR-2000; 2000US-0190076.  
 PR 18-APR-2000; 2000US-0198123.  
 PR 19-MAY-2000; 2000US-0205515.  
 PR 07-JUN-2000; 2000US-0209467.  
 PR 28-JUN-2000; 2000US-0214886.  
 PR 30-JUN-2000; 2000US-0215135.  
 PR 07-JUL-2000; 2000US-0216647.  
 PR 07-JUL-2000; 2000US-0216880.  
 PR 11-JUL-2000; 2000US-0217487.  
 PR 11-JUL-2000; 2000US-0217496.  
 PR 14-JUL-2000; 2000US-0218290.  
 PR 26-JUL-2000; 2000US-0220963.  
 PR 26-JUL-2000; 2000US-0220964.  
 PR 14-AUG-2000; 2000US-0224518.  
 PR 14-AUG-2000; 2000US-0224519.  
 PR 14-AUG-2000; 2000US-0225213.  
 PR 14-AUG-2000; 2000US-0225214.  
 PR 14-AUG-2000; 2000US-0225266.  
 PR 14-AUG-2000; 2000US-0225267.  
 PR 14-AUG-2000; 2000US-0225268.  
 PR 14-AUG-2000; 2000US-0225270.  
 PR 14-AUG-2000; 2000US-0225447.  
 PR 14-AUG-2000; 2000US-0225757.  
 PR 14-AUG-2000; 2000US-0225758.  
 PR 14-AUG-2000; 2000US-0225759.  
 PR 18-AUG-2000; 2000US-0226279.  
 PR 22-AUG-2000; 2000US-0226681.  
 PR 22-AUG-2000; 2000US-0226868.  
 PR 23-AUG-2000; 2000US-0227102.  
 PR 30-AUG-2000; 2000US-0228924.  
 PR 01-SEP-2000; 2000US-0229287.  
 PR 01-SEP-2000; 2000US-0229343.  
 PR 01-SEP-2000; 2000US-0229344.  
 PR 01-SEP-2000; 2000US-0229345.  
 PR 05-SEP-2000; 2000US-0229509.  
 PR 05-SEP-2000; 2000US-0229513.  
 PR 06-SEP-2000; 2000US-0230437.  
 PR 06-SEP-2000; 2000US-0230438.  
 PR 08-SEP-2000; 2000US-0231242.  
 PR 08-SEP-2000; 2000US-0231243.  
 PR 08-SEP-2000; 2000US-0231244.  
 PR 08-SEP-2000; 2000US-0231413.  
 PR 08-SEP-2000; 2000US-0231414.  
 PR 08-SEP-2000; 2000US-0232080.  
 PR 08-SEP-2000; 2000US-0232081.  
 PR 12-SEP-2000; 2000US-0231968.  
 PR 14-SEP-2000; 2000US-0232397.  
 PR 14-SEP-2000; 2000US-0232398.  
 PR 14-SEP-2000; 2000US-0232399.  
 PR 14-SEP-2000; 2000US-0232400.  
 PR 14-SEP-2000; 2000US-0232401.  
 PR 14-SEP-2000; 2000US-0233063.  
 PR 14-SEP-2000; 2000US-0233064.  
 PR 14-SEP-2000; 2000US-0233065.  
 PR 21-SEP-2000; 2000US-0234223.  
 PR 21-SEP-2000; 2000US-0234274.  
 PR 25-SEP-2000; 2000US-0234997.  
 PR 25-SEP-2000; 2000US-0234998.  
 PR 26-SEP-2000; 2000US-0235484.  
 PR 27-SEP-2000; 2000US-0235834.  
 PR 27-SEP-2000; 2000US-0235836.  
 PR 29-SEP-2000; 2000US-0236367.  
 PR 29-SEP-2000; 2000US-0236367.  
 PR 29-SEP-2000; 2000US-0236368.  
 PR 29-SEP-2000; 2000US-0236369.  
 PR 29-SEP-2000; 2000US-0236370.  
 PR 02-OCT-2000; 2000US-0236802.  
 PR 02-OCT-2000; 2000US-0237037.  
 PR 02-OCT-2000; 2000US-0237038.  
 PR 02-OCT-2000; 2000US-0237039.



CC relates to expression constructs comprising a promoter of the invention  
CC and DNA encoding the signal peptide, and transgenic plants and plant  
CC cells comprising the expression vector. The method of the invention is  
CC useful for the efficient production of a foreign protein in a plant, and  
CC enables the recombinantly produced protein to be easily collected. The  
CC present sequence represents cucumber XSP4.

XX Sequence 168 AA;

Query Match 45.6%; Score 41; DB 21; Length 168;  
Best Local Similarity 70.0%; Pred. No. 51;

Matches 7; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Qy 1 VGYGRDNDNR 10  
|||||  
:|:

Db 53 VGYGRDNDNR 62

RESULT 35

AAU87169  
ID AAU87169 standard; Protein; 223 AA.

XX AAU87169;

XX 05-JUN-2002 (first entry)

XX Novel central nervous system protein #79.

KM Central nervous system; CNS; autoimmune disease; rheumatoid arthritis;  
KM hyperproliferative disorder; neoplasm; cardiovascular disorder;  
KM cardiac arrest; cerebrovascular disorder; ischaemia; angiogenesis;  
KM nervous system disorder; Alzheimer's disease; AIDS; ocular disorder;  
KM acquired immunodeficiency virus; dysphagia; gastrointestinal disorder;  
KM adenocarcinoma; reproductive system disorder; testicular feminisation;  
KM endocrine disorder; diabetes; cancer; leukaemia; neovascularisation;  
KM respiratory disorder; renal disorder; kidney failure; blood disorder;  
KM myocardial infarction; wound healing; cell proliferation; skin aging;  
KM food additive; food preservative; gene therapy.

XX Homo sapiens.

XX WO200155318-A2.

XX 02-AUG-2001.

PF 17-JAN-2001; 2001MO-US01332.

XX 31-JAN-2000; 2000US-0179065.  
PR 04-FEB-2000; 2000US-0180628.  
PR 24-FEB-2000; 2000US-0184664.  
PR 02-MAR-2000; 2000US-0186350.  
PR 16-MAR-2000; 2000US-0189874.  
PR 17-MAR-2000; 2000US-0190076.  
PR 18-APR-2000; 2000US-0198123.  
PR 19-MAY-2000; 2000US-0205451.  
PR 07-JUN-2000; 2000US-0209467.  
PR 28-JUN-2000; 2000US-0214886.  
PR 30-JUN-2000; 2000US-0215135.  
PR 07-JUL-2000; 2000US-0216647.  
PR 07-JUL-2000; 2000US-0216880.  
PR 11-JUL-2000; 2000US-0217487.  
PR 11-JUL-2000; 2000US-0217496.  
PR 14-JUL-2000; 2000US-0218290.  
PR 26-JUL-2000; 2000US-0220963.  
PR 26-JUL-2000; 2000US-0220964.  
PR 14-AUG-2000; 2000US-0224518.  
PR 14-AUG-2000; 2000US-0224519.  
PR 14-AUG-2000; 2000US-0225213.  
PR 14-AUG-2000; 2000US-0225214.  
PR 14-AUG-2000; 2000US-0225266.  
PR 14-AUG-2000; 2000US-0225267.  
PR 14-AUG-2000; 2000US-0225268.  
PR 14-AUG-2000; 2000US-0225270.

PR 14-AUG-2000; 2000US-0225447.  
PR 14-AUG-2000; 2000US-0225757.  
PR 14-AUG-2000; 2000US-0225758.  
PR 14-AUG-2000; 2000US-0225759.  
PR 16-AUG-2000; 2000US-0226279.  
PR 22-AUG-2000; 2000US-0226681.  
PR 22-AUG-2000; 2000US-0226868.  
PR 23-AUG-2000; 2000US-0227182.  
PR 30-AUG-2000; 2000US-0228924.  
PR 01-SEP-2000; 2000US-0229287.  
PR 01-SEP-2000; 2000US-0229343.  
PR 01-SEP-2000; 2000US-0229344.  
PR 01-SEP-2000; 2000US-0229345.  
PR 05-SEP-2000; 2000US-0229509.  
PR 05-SEP-2000; 2000US-0229513.  
PR 06-SEP-2000; 2000US-0230437.  
PR 06-SEP-2000; 2000US-0230438.  
PR 08-SEP-2000; 2000US-0231242.  
PR 08-SEP-2000; 2000US-0231243.  
PR 08-SEP-2000; 2000US-0231244.  
PR 08-SEP-2000; 2000US-0231413.  
PR 08-SEP-2000; 2000US-0231414.  
PR 08-SEP-2000; 2000US-0232080.  
PR 08-SEP-2000; 2000US-0232081.  
PR 12-SEP-2000; 2000US-0231968.  
PR 14-SEP-2000; 2000US-0232397.  
PR 14-SEP-2000; 2000US-0232398.  
PR 14-SEP-2000; 2000US-0232399.  
PR 14-SEP-2000; 2000US-0232400.  
PR 14-SEP-2000; 2000US-0232401.  
PR 14-SEP-2000; 2000US-0233063.  
PR 14-SEP-2000; 2000US-0233064.  
PR 14-SEP-2000; 2000US-0233065.  
PR 21-SEP-2000; 2000US-0234223.  
PR 21-SEP-2000; 2000US-0234274.  
PR 25-SEP-2000; 2000US-0234997.  
PR 25-SEP-2000; 2000US-0234998.  
PR 25-SEP-2000; 2000US-0235484.  
PR 27-SEP-2000; 2000US-0235834.  
PR 27-SEP-2000; 2000US-0235836.  
PR 29-SEP-2000; 2000US-0236327.  
PR 29-SEP-2000; 2000US-0236367.  
PR 29-SEP-2000; 2000US-0236368.  
PR 29-SEP-2000; 2000US-0236369.  
PR 29-SEP-2000; 2000US-0236370.  
PR 02-OCT-2000; 2000US-0236802.  
PR 02-OCT-2000; 2000US-0237037.  
PR 02-OCT-2000; 2000US-0237038.  
PR 02-OCT-2000; 2000US-0237039.  
PR 02-OCT-2000; 2000US-0237040.  
PR 13-OCT-2000; 2000US-0239935.  
PR 13-OCT-2000; 2000US-0239937.  
PR 20-OCT-2000; 2000US-0240960.  
PR 20-OCT-2000; 2000US-0241121.  
PR 20-OCT-2000; 2000US-0241785.  
PR 20-OCT-2000; 2000US-0241786.  
PR 20-OCT-2000; 2000US-0241787.  
PR 20-OCT-2000; 2000US-0241808.  
PR 20-OCT-2000; 2000US-0241809.  
PR 20-OCT-2000; 2000US-0241826.  
PR 01-NOV-2000; 2000US-0244617.  
PR 08-NOV-2000; 2000US-0246474.  
PR 08-NOV-2000; 2000US-0246475.  
PR 08-NOV-2000; 2000US-0246476.  
PR 08-NOV-2000; 2000US-0246477.  
PR 08-NOV-2000; 2000US-0246523.  
PR 08-NOV-2000; 2000US-0246524.  
PR 08-NOV-2000; 2000US-0246525.  
PR 08-NOV-2000; 2000US-0246526.  
PR 08-NOV-2000; 2000US-0246527.  
PR 08-NOV-2000; 2000US-0246528.

PR 08-NOV-2000; 2000US-0246532.  
PR 08-NOV-2000; 2000US-0246609.  
PR 08-NOV-2000; 2000US-0246610.  
PR 08-NOV-2000; 2000US-0246611.  
PR 08-NOV-2000; 2000US-0246613.  
PR 17-NOV-2000; 2000US-0249207.  
PR 17-NOV-2000; 2000US-0249208.  
PR 17-NOV-2000; 2000US-0249209.  
PR 17-NOV-2000; 2000US-0249210.  
PR 17-NOV-2000; 2000US-0249211.  
PR 17-NOV-2000; 2000US-0249212.  
PR 17-NOV-2000; 2000US-0249213.  
PR 17-NOV-2000; 2000US-0249214.  
PR 17-NOV-2000; 2000US-0249215.  
PR 17-NOV-2000; 2000US-0249216.  
PR 17-NOV-2000; 2000US-0249217.  
PR 17-NOV-2000; 2000US-0249218.  
PR 17-NOV-2000; 2000US-0249219.  
PR 17-NOV-2000; 2000US-0249220.  
PR 17-NOV-2000; 2000US-0249221.  
PR 17-NOV-2000; 2000US-0249222.  
PR 17-NOV-2000; 2000US-0249223.  
PR 17-NOV-2000; 2000US-0249224.  
PR 17-NOV-2000; 2000US-0249225.  
PR 17-NOV-2000; 2000US-0249226.  
PR 17-NOV-2000; 2000US-0249227.  
PR 17-NOV-2000; 2000US-0249228.  
PR 17-NOV-2000; 2000US-0249229.  
PR 17-NOV-2000; 2000US-0249230.  
PR 01-DEC-2000; 2000US-0250160.  
PR 01-DEC-2000; 2000US-0250391.  
PR 05-DEC-2000; 2000US-0251030.  
PR 05-DEC-2000; 2000US-0251988.  
PR 05-DEC-2000; 2000US-0256719.  
PR 06-DEC-2000; 2000US-0251479.  
PR 08-DEC-2000; 2000US-0251856.  
PR 08-DEC-2000; 2000US-0251856.  
PR 08-DEC-2000; 2000US-0251868.  
PR 08-DEC-2000; 2000US-0251869.  
PR 08-DEC-2000; 2000US-0251989.  
PR 11-DEC-2000; 2000US-0254097.  
PR 05-JAN-2001; 2001US-0259678.  
XX  
XX (HUMA-) HUMAN GENOME SCI INC.  
XX  
PI Rosen CA, Baraeth SC, Ruben SM;  
XX WPI; 2001-581633/65.  
DR N-PSDB; ABK43499.  
XX  
XX  
PT New isolated nucleic acid encoding a protein for diagnosing,  
PT preventing, treating or ameliorating medical conditions and used as  
PT food additives or preservatives -  
XX  
XX  
XX Claim 9; SEQ ID No 687; 837bp; English.  
XX  
XX The invention describes an isolated nucleic acid molecule (I) encoding a  
XX novel central nervous system protein. (I) and polypeptides (III) encoded  
XX by (I), are used to treat a medical conditions and in diagnosis of a  
XX pathological condition. Disorders which are diagnosed or treated include  
XX autoimmune diseases e.g. rheumatoid arthritis, hyperproliferative  
XX disorders e.g. neoplasms of the breast or liver, cardiovascular disorders  
XX e.g. cardiac arrest, cerebrovascular disorders e.g. cerebral ischaemia,  
XX angiogenesis, nervous system disorders e.g. Alzheimer's disease and  
XX amyotrophic lateral sclerosis, infections caused by bacteria, viruses  
XX e.g. Acquired immunodeficiency virus (AIDS) and fungi, ocular disorders  
XX e.g. corneal infection, gastrointestinal disorders e.g. dysphagia,  
XX adenocarcinomas and irritable bowel syndrome, reproductive system  
XX disorders e.g. testicular feminisation, endocrine disorders e.g. diabetes  
XX and pituitary dwarfism, cancers and disorders at the cellular level e.g.  
XX leukaemia, disorders involving neovascularisation e.g. malignancies,  
XX respiratory disorders e.g. nonallergic rhinitis, renal disorders e.g.  
XX acute kidney failure and blood related disorders e.g. myocardial  
XX infarction. The polypeptides can also be used to aid wound healing and  
XX epithelial cell proliferation, to prevent skin aging due to sunburn, to  
XX maintain organs before transplantation, for supporting cell culture of  
XX primary tissues, to regenerate tissues and in chemotaxis. The  
XX polypeptides can also be used as a food additive or preservative to

CC increase or decrease storage capabilities, fat content, lipid, protein,  
Query Match 45.6%; Score 41; DB 22; Length 223;  
Best Local Similarity 50.0%; Pred. No. 70;  
Matches 7; Conservative 3; Mismatches 4; Indels 0; Gaps 0;  
QY 3 YGRDNDRDPGRKN 16  
: : | | | : :  
Db 84 FGKDKORDSSVRN 97  
RESULT 36  
AAU17282  
ID AAU17282 standard; Protein; 229 AA.  
XX  
AC AAU17282;  
XX  
DT 07-NOV-2001 (first entry)  
XX  
DE Novel signal transduction pathway protein, Seq ID 847.  
XX  
KW Neuroprotective; cytosolic; dermatological; immunosuppressive; tumour;  
KW antiinflammatory; anti-HIV; antibacterial; antiinflammatory; cancer;  
KW immune system disorder; rheumatoid arthritis; inflammatory condition;  
KW organ transplant rejection; infection; hepatitis C; blood disorder;  
KW sickle cell anaemia; hyperproliferative disorder; Gaucher's disease;  
KW neurodegenerative disorder; Alzheimer's disease; Parkinson's disease;  
KW chromosomal abnormality; Down syndrome; ischaemia; renal disorder;  
KW cardiovascular; respiratory; wound healing; endocrine; Addison's disease;  
KW reproductive system; gastrointestinal; liver disorder; AIDS;  
KW acquired immune deficiency syndrome.  
XX  
OS Homo sapiens.  
XX  
PN WO200154733-A1.  
XX  
PD 02-AUG-2001.  
XX  
XX  
XX 17-JAN-2001; 2001WO-US01312.  
XX  
XX 31-JAN-2000; 2000US-0179065.  
XX 04-FEB-2000; 2000US-0180628.  
XX 24-FEB-2000; 2000US-0184664.  
XX 02-MAR-2000; 2000US-0186350.  
XX 16-MAR-2000; 2000US-0189874.  
XX 17-MAR-2000; 2000US-0190076.  
XX 18-APR-2000; 2000US-0198123.  
XX 19-MAY-2000; 2000US-0205515.  
XX 07-JUN-2000; 2000US-0209467.  
XX 28-JUN-2000; 2000US-0214886.  
XX 30-JUN-2000; 2000US-0215135.  
XX 07-JUL-2000; 2000US-0216647.  
XX 07-JUL-2000; 2000US-0216880.  
XX 11-JUL-2000; 2000US-0217487.  
XX 11-JUL-2000; 2000US-0217496.  
XX 14-JUL-2000; 2000US-0218290.  
XX 26-JUL-2000; 2000US-0220963.  
XX 26-JUL-2000; 2000US-0220964.  
XX 14-AUG-2000; 2000US-0224518.  
XX 14-AUG-2000; 2000US-0224519.  
XX 14-AUG-2000; 2000US-0225213.  
XX 14-AUG-2000; 2000US-0225214.  
XX 14-AUG-2000; 2000US-0225266.  
XX 14-AUG-2000; 2000US-0225267.  
XX 14-AUG-2000; 2000US-0225268.  
XX 14-AUG-2000; 2000US-0225270.  
XX 14-AUG-2000; 2000US-0225447.  
XX 14-AUG-2000; 2000US-0225757.  
XX 14-AUG-2000; 2000US-0225758.  
XX 14-AUG-2000; 2000US-0225759.  
XX 18-AUG-2000; 2000US-0226279.  
XX 22-AUG-2000; 2000US-0226681.  
XX 22-AUG-2000; 2000US-0226681.



Db 90 FGKDKORDSSVRN 103

:|:| | | | :|

## RESULT 37

AAE15538 ID AAE15538 standard; Protein; 310 AA.

XX AAE15538;

DT 12-MAR-2002 (first entry)

DE Mus musculus U2 snRNP auxiliary factor protein.

XX Environmental tolerance; salt toxicity; transgenic plant; osmotic;

KM cold; drought; freezing stress; messenger RNA processing; plant growth;

KW U2 snRNP auxiliary factor; mouse.

XX Mus musculus.

PN WO200181599-A2.

XX 01-NOV-2001.

PF 19-APR-2001; 2001WO-EP04479.

XX 19-APR-2000; 2000ES-0001102.

XX (UVVA-) UNIV VALENCIA POLITECNICA.

PI Vicente Meana O, Roldan Medina M, Serrano Salom R;

PI Forment Millet JJ, Naranjo Olivero MA;

DR WPI: 2002-066441/09.

DR N-PSDB; AAD24770.

XX Claim 26; Fig 5; 84pp; English.

XX The invention relates to a method for enhancing environmental stress  
XX tolerance or salt toxicity in eukaryotic cells and organisms by  
XX manipulating the process of processing messenger RNA precursors. The  
XX invention also relates to isolated nucleic acids (Beta vulgaris genes  
XX that enhance salt tolerance, Mus musculus gene that encode small unit  
XX of U2 snRNP auxiliary factor protein) encoding a protein or  
XX immunologically active and/or functional fragment of it, involved in  
XX salt tolerance. The method is useful for enhancing stress tolerance  
XX comprising tolerance against osmotic, drought, cold or freezing stress  
XX and salt toxicity in cells and organisms. The isolated nucleic acid in  
XX an expressible format is useful in the production of transgenic or  
XX altered plants, plant cells or plant tissues. The nucleic acids are also  
XX useful for enhancing or altering stress tolerance in a plant, for  
XX increasing the plant yield and for stimulating plant growth, which can  
XX be in any part of the plant, such as root, leaf or seed. The present  
XX sequence is mouse U2 snRNP auxiliary factor protein capable of  
XX enhancing salt tolerance.

XX Sequence 310 AA;

Query March 45 6%; Score 41; DB 23; Length 310;

Best Local Similarity 70.0%; Pred. No. 1e+02; Mismatches 0; Gaps 0;

Matches 7; Conservative 2; Indels 0;

Db 215 GYGRDDDDRN 224

RESULT 38

AA36931 ID AA36931 standard; Protein; 376 AA.

XX AA36931;

DT 18-OCT-2000 (first entry)

DE Arabidopsis thaliana protein fragment SEQ ID NO: 45328.

XX Protein identification; signal transduction pathway; metabolic pathway;

KM hybridisation assay; genetic mapping; gene expression control; promoter;

KW termination sequence.

XX Arabidopsis thaliana.

EN EP1033405-A2.

PD 06-SEP-2000.

XX 25-FEB-2000; 2000EP-0301439.

XX 25-FEB-1999; 99US-0121825.

PR 05-MAR-1999; 99US-0123180.

PR 09-MAR-1999; 99US-0123548.

PR 23-MAR-1999; 99US-0125788.

PR 25-MAR-1999; 99US-0126264.

PR 29-MAR-1999; 99US-0126785.

PR 01-APR-1999; 99US-0127462.

PR 06-APR-1999; 99US-0128234.

PR 08-APR-1999; 99US-0128714.

PR 16-APR-1999; 99US-0129845.

PR 19-APR-1999; 99US-0130077.

PR 21-APR-1999; 99US-0130449.

PR 23-APR-1999; 99US-0130510.

PR 28-APR-1999; 99US-0130891.

PR 30-APR-1999; 99US-0131449.

PR 30-APR-1999; 99US-0132048.

PR 04-MAY-1999; 99US-0132407.

PR 05-MAY-1999; 99US-0132484.

PR 06-MAY-1999; 99US-0132485.

PR 06-MAY-1999; 99US-0132486.

PR 07-MAY-1999; 99US-0132487.

PR 11-MAY-1999; 99US-0132863.

PR 14-MAY-1999; 99US-0134256.

PR 14-MAY-1999; 99US-0134219.

PR 14-MAY-1999; 99US-0134221.

PR 18-MAY-1999; 99US-0134370.

PR 20-MAY-1999; 99US-0134968.

PR 21-MAY-1999; 99US-0135124.

PR 24-MAY-1999; 99US-0135353.

PR 25-MAY-1999; 99US-0136021.

PR 27-MAY-1999; 99US-0136392.

PR 28-MAY-1999; 99US-0137682.

PR 01-JUN-1999; 99US-0137222.

PR 03-JUN-1999; 99US-0137528.

PR 04-JUN-1999; 99US-0137502.

PR 07-JUN-1999; 99US-0137724.

PR 08-JUN-1999; 99US-0138094.

PR 10-JUN-1999; 99US-0138540.

PR 10-JUN-1999; 99US-0138847.

PR 14-JUN-1999; 99US-0139119.

PR 16-JUN-1999; 99US-0139452.

PR 16-JUN-1999; 99US-0139453.

PR 17-JUN-1999; 99US-0139454.

PR 18-JUN-1999; 99US-0139454.

PR 18-JUN-1999; 99US-0139455.

PR 18-JUN-1999; 99US-0139456.

PR 18-JUN-1999; 99US-0139457.

PR 18-JUN-1999; 99US-0139458.

PR 18-JUN-1999; 99US-0139459.

PR 18-JUN-1999; 99US-0139460.  
PR 18-JUN-1999; 99US-0139461.  
PR 18-JUN-1999; 99US-0139462.  
PR 18-JUN-1999; 99US-0139463.  
PR 18-JUN-1999; 99US-0139750.  
PR 18-JUN-1999; 99US-0139763.  
PR 21-JUN-1999; 99US-0139817.  
PR 22-JUN-1999; 99US-0139899.  
PR 23-JUN-1999; 99US-0140353.  
PR 23-JUN-1999; 99US-0140354.  
PR 24-JUN-1999; 99US-0140695.  
PR 28-JUN-1999; 99US-0140823.  
PR 29-JUN-1999; 99US-0140991.  
PR 30-JUN-1999; 99US-0141287.  
PR 01-JUL-1999; 99US-0141842.  
PR 01-JUL-1999; 99US-0142154.  
PR 02-JUL-1999; 99US-0142055.  
PR 06-JUL-1999; 99US-0142390.  
PR 08-JUL-1999; 99US-0142803.  
PR 09-JUL-1999; 99US-0142920.  
PR 12-JUL-1999; 99US-0142977.  
PR 13-JUL-1999; 99US-0143542.  
PR 14-JUL-1999; 99US-0143624.  
PR 15-JUL-1999; 99US-0144005.  
PR 16-JUL-1999; 99US-0144085.  
PR 16-JUL-1999; 99US-0144086.  
PR 19-JUL-1999; 99US-0144325.  
PR 19-JUL-1999; 99US-0144331.  
PR 19-JUL-1999; 99US-0144332.  
PR 19-JUL-1999; 99US-0144333.  
PR 19-JUL-1999; 99US-0144334.  
PR 19-JUL-1999; 99US-0144335.  
PR 20-JUL-1999; 99US-0144352.  
PR 20-JUL-1999; 99US-0144632.  
PR 20-JUL-1999; 99US-0144884.  
PR 21-JUL-1999; 99US-0144814.  
PR 21-JUL-1999; 99US-0145086.  
PR 21-JUL-1999; 99US-0145087.  
PR 22-JUL-1999; 99US-0145088.  
PR 22-JUL-1999; 99US-0145089.  
PR 22-JUL-1999; 99US-0145192.  
PR 23-JUL-1999; 99US-0145218.  
PR 23-JUL-1999; 99US-0145219.  
PR 23-JUL-1999; 99US-0145224.  
PR 26-JUL-1999; 99US-0145276.  
PR 27-JUL-1999; 99US-0145913.  
PR 27-JUL-1999; 99US-0145918.  
PR 27-JUL-1999; 99US-0145919.  
PR 28-JUL-1999; 99US-0145951.  
PR 02-AUG-1999; 99US-0146386.  
PR 02-AUG-1999; 99US-0146388.  
PR 02-AUG-1999; 99US-0146389.  
PR 03-AUG-1999; 99US-0147038.  
PR 04-AUG-1999; 99US-0147204.  
PR 04-AUG-1999; 99US-0147302.  
PR 05-AUG-1999; 99US-0147192.  
PR 05-AUG-1999; 99US-0147260.  
PR 06-AUG-1999; 99US-0147303.  
PR 06-AUG-1999; 99US-0147416.  
PR 09-AUG-1999; 99US-0147493.  
PR 09-AUG-1999; 99US-0147935.  
PR 10-AUG-1999; 99US-0148171.  
PR 11-AUG-1999; 99US-0148319.  
PR 12-AUG-1999; 99US-0148341.  
PR 13-AUG-1999; 99US-0148565.  
PR 13-AUG-1999; 99US-0148684.  
PR 16-AUG-1999; 99US-0149368.  
PR 17-AUG-1999; 99US-0149175.  
PR 18-AUG-1999; 99US-0149426.  
PR 20-AUG-1999; 99US-0149722.  
PR 20-AUG-1999; 99US-0149723.  
PR 20-AUG-1999; 99US-0149929.

PR 23-AUG-1999; 99US-0149902.  
PR 23-AUG-1999; 99US-0149930.  
PR 25-AUG-1999; 99US-0150566.  
PR 26-AUG-1999; 99US-0150884.  
PR 27-AUG-1999; 99US-0151065.  
PR 27-AUG-1999; 99US-0151066.  
PR 27-AUG-1999; 99US-0151066.  
PR 30-AUG-1999; 99US-0151080.  
PR 31-AUG-1999; 99US-0151303.  
PR 01-SEP-1999; 99US-0151930.  
PR 07-SEP-1999; 99US-0152363.  
PR 10-SEP-1999; 99US-0153070.  
PR 13-SEP-1999; 99US-0153758.  
PR 15-SEP-1999; 99US-0154039.  
PR 16-SEP-1999; 99US-0154039.  
PR 20-SEP-1999; 99US-0154779.  
PR 22-SEP-1999; 99US-0155139.  
PR 23-SEP-1999; 99US-0155486.  
PR 24-SEP-1999; 99US-0155659.  
PR 28-SEP-1999; 99US-0156458.  
PR 29-SEP-1999; 99US-0156596.  
PR 04-OCT-1999; 99US-0157117.  
PR 05-OCT-1999; 99US-0157753.  
PR 06-OCT-1999; 99US-0157865.  
PR 07-OCT-1999; 99US-0158029.  
PR 08-OCT-1999; 99US-0158232.  
PR 12-OCT-1999; 99US-0158369.  
PR 13-OCT-1999; 99US-0159293.  
PR 13-OCT-1999; 99US-0159294.  
PR 13-OCT-1999; 99US-0159295.  
PR 13-OCT-1999; 99US-0159295.  
PR 14-OCT-1999; 99US-0159329.  
PR 14-OCT-1999; 99US-0159330.  
PR 14-OCT-1999; 99US-0159331.  
PR 14-OCT-1999; 99US-0159637.  
PR 14-OCT-1999; 99US-0159638.  
PR 18-OCT-1999; 99US-0159584.  
PR 21-OCT-1999; 99US-0160741.  
PR 21-OCT-1999; 99US-0160767.  
PR 21-OCT-1999; 99US-0160768.  
PR 21-OCT-1999; 99US-0160770.  
PR 21-OCT-1999; 99US-0160814.  
PR 21-OCT-1999; 99US-0160815.  
PR 22-OCT-1999; 99US-0160980.  
PR 22-OCT-1999; 99US-0160981.  
PR 22-OCT-1999; 99US-0160989.  
PR 25-OCT-1999; 99US-0161404.  
PR 25-OCT-1999; 99US-0161405.  
PR 25-OCT-1999; 99US-0161406.  
PR 26-OCT-1999; 99US-0161359.  
PR 26-OCT-1999; 99US-0161360.  
PR 26-OCT-1999; 99US-0161361.  
PR 28-OCT-1999; 99US-0161920.  
PR 28-OCT-1999; 99US-0161992.  
PR 28-OCT-1999; 99US-0161993.  
PR 29-OCT-1999; 99US-0162142.

Query Match 45.6%; Score 41; DB 21; Length 376;  
Best Local Similarity 62.5%; Pred. No. 1.3e+02; Indels 4; Gaps 1;  
Matches 10; Conservative 1; Mismatches 1;

Oy 4 GRD----DNDRPSRK 15  
Db 274 GRDRTSDHRRSRK 269

RESULT 39  
AAG36930  
ID AAG36930 standard; Protein: 415 AA.

AC AAG36930;  
XX  
DT 18-OCT-2000 (first entry)  
XX

DE Arabidopsis thaliana protein fragment SEQ ID NO: 45327.  
XX  
XX Protein identification; signal transduction pathway; metabolic pathway;  
KM hybridisation assay; genetic mapping; gene expression control; promoter;  
KM termination sequence.  
XX  
OS Arabidopsis thaliana.  
XX  
XX EP1033405-A2.  
XX  
PD 06-SEP-2000.  
XX  
PF 25-FEB-2000; 2000EP-0301439.  
XX  
XX 25-FEB-1999; 99US-0121825.  
PR 05-MAR-1999; 99US-0123180.  
PR 09-MAR-1999; 99US-0123548.  
PR 23-MAR-1999; 99US-0125788.  
PR 25-MAR-1999; 99US-0126264.  
PR 29-MAR-1999; 99US-0126785.  
PR 01-APR-1999; 99US-0127462.  
PR 06-APR-1999; 99US-0128234.  
PR 08-APR-1999; 99US-0128714.  
PR 16-APR-1999; 99US-0129845.  
PR 19-APR-1999; 99US-0130077.  
PR 21-APR-1999; 99US-0130449.  
PR 23-APR-1999; 99US-0130510.  
PR 28-APR-1999; 99US-0130891.  
PR 30-APR-1999; 99US-0131449.  
PR 30-APR-1999; 99US-0132048.  
PR 30-APR-1999; 99US-0132407.  
PR 04-MAY-1999; 99US-0132484.  
PR 05-MAY-1999; 99US-0132485.  
PR 06-MAY-1999; 99US-0132486.  
PR 07-MAY-1999; 99US-0132487.  
PR 11-MAY-1999; 99US-0132863.  
PR 14-MAY-1999; 99US-0134256.  
PR 14-MAY-1999; 99US-0134218.  
PR 14-MAY-1999; 99US-0134219.  
PR 14-MAY-1999; 99US-0134370.  
PR 14-MAY-1999; 99US-0134370.  
PR 18-MAY-1999; 99US-0134768.  
PR 19-MAY-1999; 99US-0134941.  
PR 20-MAY-1999; 99US-0135124.  
PR 21-MAY-1999; 99US-0135353.  
PR 24-MAY-1999; 99US-0135629.  
PR 25-MAY-1999; 99US-0136021.  
PR 27-MAY-1999; 99US-0136392.  
PR 28-MAY-1999; 99US-0136782.  
PR 01-JUN-1999; 99US-0137222.  
PR 03-JUN-1999; 99US-0137528.  
PR 04-JUN-1999; 99US-0137502.  
PR 07-JUN-1999; 99US-0137724.  
PR 08-JUN-1999; 99US-0138094.  
PR 10-JUN-1999; 99US-0138540.  
PR 10-JUN-1999; 99US-0138847.  
PR 14-JUN-1999; 99US-0139119.  
PR 16-JUN-1999; 99US-0139452.  
PR 17-JUN-1999; 99US-0139453.  
PR 17-JUN-1999; 99US-0139452.  
PR 18-JUN-1999; 99US-0139455.  
PR 18-JUN-1999; 99US-0139456.  
PR 18-JUN-1999; 99US-0139457.  
PR 18-JUN-1999; 99US-0139458.  
PR 18-JUN-1999; 99US-0139459.  
PR 18-JUN-1999; 99US-0139460.  
PR 18-JUN-1999; 99US-0139461.  
PR 18-JUN-1999; 99US-0139462.  
PR 18-JUN-1999; 99US-0139463.  
PR 18-JUN-1999; 99US-0139750.  
PR 18-JUN-1999; 99US-0139763.  
PR 21-JUN-1999; 99US-0139817.

PR 22-JUN-1999; 99US-0139899.  
PR 23-JUN-1999; 99US-0140353.  
PR 23-JUN-1999; 99US-0140354.  
PR 24-JUN-1999; 99US-0140695.  
PR 28-JUN-1999; 99US-0140823.  
PR 29-JUN-1999; 99US-0140991.  
PR 30-JUN-1999; 99US-0141287.  
PR 01-JUL-1999; 99US-0141842.  
PR 02-JUL-1999; 99US-0142154.  
PR 06-JUL-1999; 99US-0142055.  
PR 06-JUL-1999; 99US-0142390.  
PR 08-JUL-1999; 99US-0142803.  
PR 09-JUL-1999; 99US-0142920.  
PR 12-JUL-1999; 99US-0142977.  
PR 13-JUL-1999; 99US-0143542.  
PR 14-JUL-1999; 99US-0143624.  
PR 15-JUL-1999; 99US-0144005.  
PR 16-JUL-1999; 99US-0144085.  
PR 16-JUL-1999; 99US-0144086.  
PR 19-JUL-1999; 99US-0144325.  
PR 19-JUL-1999; 99US-0144331.  
PR 19-JUL-1999; 99US-0144332.  
PR 19-JUL-1999; 99US-0144333.  
PR 19-JUL-1999; 99US-0144334.  
PR 19-JUL-1999; 99US-0144335.  
PR 20-JUL-1999; 99US-0144352.  
PR 20-JUL-1999; 99US-0144632.  
PR 20-JUL-1999; 99US-0144884.  
PR 21-JUL-1999; 99US-0144814.  
PR 21-JUL-1999; 99US-0145086.  
PR 21-JUL-1999; 99US-0145088.  
PR 22-JUL-1999; 99US-0145085.  
PR 22-JUL-1999; 99US-0145087.  
PR 22-JUL-1999; 99US-0145089.  
PR 22-JUL-1999; 99US-0145192.  
PR 23-JUL-1999; 99US-0145145.  
PR 23-JUL-1999; 99US-0145218.  
PR 23-JUL-1999; 99US-0145224.  
PR 26-JUL-1999; 99US-0145276.  
PR 27-JUL-1999; 99US-0145913.  
PR 27-JUL-1999; 99US-0145918.  
PR 27-JUL-1999; 99US-0145919.  
PR 28-JUL-1999; 99US-0145951.  
PR 02-AUG-1999; 99US-0146386.  
PR 02-AUG-1999; 99US-0146388.  
PR 02-AUG-1999; 99US-0146389.  
PR 03-AUG-1999; 99US-0147038.  
PR 04-AUG-1999; 99US-0147204.  
PR 04-AUG-1999; 99US-0147302.  
PR 05-AUG-1999; 99US-0147192.  
PR 05-AUG-1999; 99US-0147260.  
PR 06-AUG-1999; 99US-0147303.  
PR 06-AUG-1999; 99US-0147416.  
PR 09-AUG-1999; 99US-0147493.  
PR 09-AUG-1999; 99US-0147935.  
PR 10-AUG-1999; 99US-0148171.  
PR 11-AUG-1999; 99US-0148319.  
PR 12-AUG-1999; 99US-0148341.  
PR 13-AUG-1999; 99US-0148565.  
PR 13-AUG-1999; 99US-0148684.  
PR 16-AUG-1999; 99US-0149368.  
PR 17-AUG-1999; 99US-0149175.  
PR 18-AUG-1999; 99US-0149426.  
PR 18-AUG-1999; 99US-0149722.  
PR 20-AUG-1999; 99US-0149723.  
PR 20-AUG-1999; 99US-0149929.  
PR 20-AUG-1999; 99US-0149930.  
PR 23-AUG-1999; 99US-0149930.  
PR 23-AUG-1999; 99US-0149930.  
PR 25-AUG-1999; 99US-0150566.  
PR 26-AUG-1999; 99US-0150884.  
PR 27-AUG-1999; 99US-0151065.  
PR 27-AUG-1999; 99US-0151066.  
PR 27-AUG-1999; 99US-0151080.





PR 01-JUL-1999; 99US-0141842.  
 PR 01-JUL-1999; 99US-0142154.  
 PR 02-JUL-1999; 99US-0142055.  
 PR 06-JUL-1999; 99US-0142390.  
 PR 08-JUL-1999; 99US-0142803.  
 PR 09-JUL-1999; 99US-0142920.  
 PR 12-JUL-1999; 99US-0142977.  
 PR 13-JUL-1999; 99US-0143542.  
 PR 14-JUL-1999; 99US-0143624.  
 PR 15-JUL-1999; 99US-0144005.  
 PR 16-JUL-1999; 99US-0144085.  
 PR 16-JUL-1999; 99US-0144086.  
 PR 19-JUL-1999; 99US-0144325.  
 PR 19-JUL-1999; 99US-0144331.  
 PR 19-JUL-1999; 99US-0144332.  
 PR 19-JUL-1999; 99US-0144333.  
 PR 19-JUL-1999; 99US-0144334.  
 PR 19-JUL-1999; 99US-0144335.  
 PR 20-JUL-1999; 99US-0144352.  
 PR 20-JUL-1999; 99US-0144632.  
 PR 20-JUL-1999; 99US-0144884.  
 PR 21-JUL-1999; 99US-0144814.  
 PR 21-JUL-1999; 99US-0145086.  
 PR 21-JUL-1999; 99US-0145088.  
 PR 22-JUL-1999; 99US-0145088.  
 PR 22-JUL-1999; 99US-0145087.  
 PR 22-JUL-1999; 99US-0145089.  
 PR 22-JUL-1999; 99US-0145192.  
 PR 23-JUL-1999; 99US-0145145.  
 PR 23-JUL-1999; 99US-0145218.  
 PR 23-JUL-1999; 99US-0145224.  
 PR 26-JUL-1999; 99US-0145276.  
 PR 27-JUL-1999; 99US-0145913.  
 PR 27-JUL-1999; 99US-0145918.  
 PR 27-JUL-1999; 99US-0145919.  
 PR 28-JUL-1999; 99US-0145951.  
 PR 02-AUG-1999; 99US-0146386.  
 PR 02-AUG-1999; 99US-0146388.  
 PR 02-AUG-1999; 99US-0146389.  
 PR 03-AUG-1999; 99US-0147038.  
 PR 04-AUG-1999; 99US-0147204.  
 PR 04-AUG-1999; 99US-0147302.  
 PR 05-AUG-1999; 99US-0147192.  
 PR 05-AUG-1999; 99US-0147260.  
 PR 06-AUG-1999; 99US-0147303.  
 PR 06-AUG-1999; 99US-0147416.  
 PR 09-AUG-1999; 99US-0147493.  
 PR 09-AUG-1999; 99US-0147935.  
 PR 10-AUG-1999; 99US-0148171.  
 PR 11-AUG-1999; 99US-0148319.  
 PR 12-AUG-1999; 99US-0148341.  
 PR 13-AUG-1999; 99US-0148565.  
 PR 13-AUG-1999; 99US-0148684.  
 PR 16-AUG-1999; 99US-0149368.  
 PR 17-AUG-1999; 99US-0149175.  
 PR 18-AUG-1999; 99US-0149426.  
 PR 20-AUG-1999; 99US-0149722.  
 PR 20-AUG-1999; 99US-0149723.  
 PR 20-AUG-1999; 99US-0149929.  
 PR 23-AUG-1999; 99US-0149902.  
 PR 23-AUG-1999; 99US-0149930.  
 PR 25-AUG-1999; 99US-0150566.  
 PR 26-AUG-1999; 99US-0150884.  
 PR 27-AUG-1999; 99US-0151065.  
 PR 27-AUG-1999; 99US-0151066.  
 PR 27-AUG-1999; 99US-0151080.  
 PR 30-AUG-1999; 99US-0151303.  
 PR 31-AUG-1999; 99US-0151438.  
 PR 01-SEP-1999; 99US-0151930.  
 PR 07-SEP-1999; 99US-0152363.  
 PR 10-SEP-1999; 99US-0153070.  
 PR 13-SEP-1999; 99US-0153758.  
 PR 15-SEP-1999; 99US-0154018.

PR 16-SEP-1999; 99US-0154039.  
 PR 20-SEP-1999; 99US-0154779.  
 PR 22-SEP-1999; 99US-0155139.  
 PR 23-SEP-1999; 99US-0155486.  
 PR 24-SEP-1999; 99US-0155659.  
 PR 28-SEP-1999; 99US-0156458.  
 PR 29-SEP-1999; 99US-0156596.  
 PR 04-OCT-1999; 99US-0157117.  
 PR 05-OCT-1999; 99US-0157753.  
 PR 06-OCT-1999; 99US-0157865.  
 PR 07-OCT-1999; 99US-0158029.  
 PR 08-OCT-1999; 99US-0158232.  
 PR 12-OCT-1999; 99US-0158369.  
 PR 13-OCT-1999; 99US-0159293.  
 PR 13-OCT-1999; 99US-0159294.  
 PR 13-OCT-1999; 99US-0159295.  
 PR 14-OCT-1999; 99US-0159329.  
 PR 14-OCT-1999; 99US-0159330.  
 PR 14-OCT-1999; 99US-0159331.  
 PR 14-OCT-1999; 99US-0159637.  
 PR 14-OCT-1999; 99US-0159638.  
 PR 18-OCT-1999; 99US-0159584.  
 PR 21-OCT-1999; 99US-0160741.  
 PR 21-OCT-1999; 99US-0160767.  
 PR 21-OCT-1999; 99US-0160768.  
 PR 21-OCT-1999; 99US-0160770.  
 PR 21-OCT-1999; 99US-0160814.  
 PR 21-OCT-1999; 99US-0160815.  
 PR 22-OCT-1999; 99US-0160980.  
 PR 22-OCT-1999; 99US-0160981.  
 PR 22-OCT-1999; 99US-0160981.  
 PR 22-OCT-1999; 99US-0160981.  
 PR 25-OCT-1999; 99US-0161404.  
 PR 25-OCT-1999; 99US-0161405.  
 PR 25-OCT-1999; 99US-0161406.  
 PR 25-OCT-1999; 99US-0161406.  
 PR 26-OCT-1999; 99US-0161359.  
 PR 26-OCT-1999; 99US-0161360.  
 PR 26-OCT-1999; 99US-0161361.  
 PR 28-OCT-1999; 99US-0161920.  
 PR 28-OCT-1999; 99US-0161992.  
 PR 28-OCT-1999; 99US-0161993.  
 PR 29-OCT-1999; 99US-0162142.

Query Match 45.6%; Score 41; DB 21; Length 448;  
 Best Local Similarity 62.5%; Pred. No. 1.5e+02;  
 Matches 10; Conservative 1; Mismatches 1; Indels 4; Gaps 1;  
 QY 4 GRD----DNDPDRK 15  
 DB 346 GRDRTSRDHRDRSRK 361

Search completed: April 9, 2003, 12:55:29  
 Job time : 15.377 secs

**This Page is Inserted by IFW Indexing and Scanning  
Operations and is not part of the Official Record**

**BEST AVAILABLE IMAGES**

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images include but are not limited to the items checked:

- ☐ BLACK BORDERS
- ☒ IMAGE CUT OFF AT TOP, BOTTOM OR SIDES
- ☐ FADED TEXT OR DRAWING
- ☐ BLURRED OR ILLEGIBLE TEXT OR DRAWING
- ☐ SKEWED/SLANTED IMAGES
- ☐ COLOR OR BLACK AND WHITE PHOTOGRAPHS
- ☐ GRAY SCALE DOCUMENTS
- ☐ LINES OR MARKS ON ORIGINAL DOCUMENT
- ☒ REFERENCE(S) OR EXHIBIT(S) SUBMITTED ARE POOR QUALITY
- ☐ OTHER: \_\_\_\_\_

**IMAGES ARE BEST AVAILABLE COPY.**

**As rescanning these documents will not correct the image problems checked, please do not report these problems to the IFW Image Problem Mailbox.**

**THIS PAGE BLANK (USPTO)**